

nature

THE INTERNATIONAL WEEKLY JOURNAL OF SCIENCE

London, UK

Thursday 3 September



18°

06 16

19 43

↑ 19°

↓ 9°

64%

1026 hPa

Fri

18°

9

Sat

19°

10

Sun

20°

12

Mon

21°

11

Tue

23°

10

Wed

22°

9

FORECAST NEWS

Can numerical weather prediction keep on getting better? **PAGE 47**



WEARABLE TECHNOLOGY

FAD OR FUTURE?

Security and data capacity are key hurdles to devices

PAGE 22



REPRODUCIBILITY

SOUND SCIENCE

Make funding dependent on good institutional practice

PAGE 25



SUPERCONDUCTIVITY

WARMING TO THE TASK

Hydrogen sulfide breaks the high- T_c record

PAGES 40 & 73

NATURE.COM/NATURE

3 September 2015 £10

Vol. 525, No. 7567



9 770028 083095

36

THIS WEEK

EDITORIALS

WORLD VIEW Nations are right to reject growth of genetically modified crops **p.7**

BOAR SCORE Pig history shows regular wild breeding events **p.8**



OBESITY Fat-burning genetic switch changes tissue fate **p.9**

Personal responsibility

The US Precision Medicine Initiative needs to tread carefully when revealing health and genetic data to participants.

When Stephen Damiani's one-year-old son Massimo suddenly lost his ability to crawl and developed other problems in 2009, doctors could not diagnose him and told his parents that he was unlikely to live long. But Stephen, who is not a scientist, had his family's genomes sequenced and worked with geneticist Ryan Taft at the University of Queensland in Australia to identify a mutated gene.

Taft linked the gene to a class of neurodegenerative disorders involving the myelin sheath, which protects neurons. The discovery allowed Massimo to be treated with therapies for related conditions, and it helped a dozen or so other families worldwide who realized that their children had the same disorder.

Such commendable initiative and diligence is a testament to the power and promise of precision medicine — therapies targeted to individuals. But unlike rare conditions such as Massimo's, the overwhelming majority of genetic and environmental factors linked to common diseases contribute only slightly to disease risk. Hundreds of genes are probably involved in depression or breast cancer, for instance.

Researchers and ethicists have spent decades struggling with the question of how much data to release to patients. Many argue that revealing information about disease risk to individuals is unnecessary and irresponsible, owing to the potential for misinterpretation. Stoking people's fears in this way could lead to expensive, unwarranted and invasive medical tests. And such information could perpetuate the idea of genes as destiny — a low genetic risk for heart disease, for instance, could be used as an excuse to eschew a healthy diet or exercise. The darker side of the argument, which scientists hesitate to state publicly, is the worry that releasing data into the public sphere too soon could scupper their chances for publication.

Historically, most clinical trials have not returned individuals' information. But recent years have seen a move towards openness, from WikiLeaks to open-access publication. The attitude that a select few should control others' data is increasingly seen as paternalistic.

Treading this careful line between professional responsibility and transparency is the US Precision Medicine Initiative. Backed by President Barack Obama, who announced it in his January State of the Union address, the project aims to collect genetic, medical, physiological and environmental data from 1 million people and follow them for decades in an attempt to link these factors with health outcomes. Its planning committee, which is expected to propose a design for the project this month, seems to be undecided on the issue: the argument came up repeatedly and heatedly at a July committee workshop (see page 16).

The simplest but most restrictive approach is to inform people about findings only once they are discovered. Yet it seems incongruous to withhold health data when decisions such as organ donation are based on the idea that the body is the person's legal possession. A better solution would be withholding data by default, but releasing them if participants request it. Ideally, that release should occur only with guidance on how to interpret the information and alongside counselling on its significance.

Such a system would stretch the Precision Medicine Initiative's underwhelming US\$215-million budget, and could burden researchers who are searching for broader trends. Not every clinical study should be expected to take such an approach. But the initiative has tried to build its brand as an atypical, egalitarian study in which participants are partners with researchers. With proper cautions in place, access on request could demonstrate how to make good on that promise. ■

Parched California

Drought highlights the state's lack of an ecological strategy.

For an unassuming little fish, the delta smelt (*Hypomesus transpacificus*) has received outsize attention. In the sprawling waterways of the Sacramento–San Joaquin river delta, which channel precious water throughout northern California, the smelt has served as an environmental sentry. When its numbers plummet, water managers flood the delta with fresh water, to the outrage of farmers who would rather have it nourishing their crops.

Yet the drought may finally do for the smelt. As California looks to enter its fifth year of drought, officials face difficult choices on how to manage water over the long term. So far, thanks to resilience built into

the water system in past years, Californians are weathering the shortage remarkably well (see Amir AghaKouchak *et al. Nature* **524**, 409–411; 2015). Cities have opted to control their love of lush lawns, and farmers have shifted to efficient irrigation and other water-saving measures.

But how long can the Golden State's lustre last? Two new reports (see go.nature.com/jpze97 and go.nature.com/okxrdo) highlight possible futures should the drought continue. And the outlook is not always that promising. Water managers cannot simply hope for a rainy winter, perhaps prompted by El Niño. Farmers will still pump groundwater for California's US\$46-billion agricultural industry, so water tables will continue to drop. More at risk are California's iconic ecosystems, from towering redwood trees to rivers teeming with salmon and trout. Wildlife managers have arranged to keep the most crucial wetlands damp for bird visits, and forestry managers extinguish wildfires as soon as they start. But such piecemeal approaches must be turned into a long-term strategy, much as farmers and urban planners have already done for their thirsty constituents.

Otherwise, the delta smelt may vanish for good. ■



Rejection of GM crops is not a failure for science

Governments maintaining their antipathy for transgenic crops are sensibly balancing public consent with scientific evidence, says Colin Macilwain.

Last week, Reuters reported that Germany is set to continue its moratorium on the cultivation of genetically modified (GM) crops. The decision will doubtless meet a well-orchestrated barrage of criticism. When the Scottish government made the same call last month, its decision was roundly condemned by plant biologists and scientific leaders such as Anne Glover, former chief scientific adviser to the president of the European Commission. Critics portray the ban as an affront to science and to the idea that regulation should be based on evidence.

I'm a big fan of the scientific method. You won't find me sitting in an Airbus 320 thanking the Lord for keeping the aircraft aloft. I happily attribute its successful flight to the scientists and engineers who mastered fluid dynamics. I also support the general principle of evidence-based policy.

Yet I'm relaxed about the pending decisions of Scotland, Germany, France, Italy and others to stand up to corporate pressure and keep GM crop technology out of the European countryside. I await with interest England's response to the deal that the European Union made last December that allows its member states to make their own choices on licensing GM crops.

Whatever these nations decide, the stakes are not as high as they once were. When the United States started to license GM soya beans and maize (corn) 20 years ago, many crop producers thought that global acceptance of the technology would rest heavily on European acceptance. That is probably no longer true. The global acreage of GM crops has grown consistently without broad acceptance from Europe. It is now topping out.

Last year, it grew by only around 3%, according to industry figures, to 181 million hectares — a little more than one-tenth of the 1.5 billion hectares of land that the United Nations estimates to be under crop cultivation.

Five-sixths of that GM acreage is in the Americas. The rest consists mostly of non-food crops (mainly cotton) grown in India and China. Little of the harvest is in nations that need improved yields to feed themselves. Twenty years in, the GM strains currently under cultivation are still best suited to the needs of large-scale industrial farmers who can afford the seeds and inputs that accompany them. Whatever Europe decides, the rest of the world isn't waiting to follow suit.

And this time, Europe's debate about GM crop cultivation isn't really over GM crops themselves, but over how nations should assess and manage risk. When Europe turned its back on GM crops 15 years ago, the pro-GM lobby warned that this signalled a continent in crisis, one unwilling to embrace the future. But there has been scant indication since then that Europe is technology-averse. It has not slowed itself down or tied itself

up by rejecting nanotechnology-based wound-dressings or mobile phones, of which it was the world's fastest adopter.

Despite the GM episode, evidence-based policy is alive and kicking in Europe. But good risk management involves early communication with the public and the careful weighing of many factors, not just scientific risk assessment. In general, however, industry — which usually holds most of the relevant data — favours scientific risk assessment as the be-all and end-all of regulation (see *Nature* 508, 289; 2014). Environmentalists — even gentle ones, such as the European Commission and former US vice-president Al Gore — prefer the precautionary principle, which places the burden of proof on the innovator.

In practice, all governments have to walk a line between the two.

But where to draw that line? In Europe, especially in countries that value the provenance of food, much of the general public doesn't want GM foods. The jury, too, remains out on their ecological impacts (see *Nature* 497, 24–26; 2013). Should they nonetheless be grown because the data say that they're safe to eat? Call me naive, but given the threadbare state of our democracy, it doesn't do to override public concern in that way.

In the United States, the key regulatory decisions were made in 1995, with scant public input. They clicked in place on the basis of 'substantial equivalence', which holds that GM foods are substantially the same as their component parts.

Substantial equivalence was the original sin that undermined public confidence in GM technology, and advocates have been over-compensating for it ever since. Genetic modification is a

blockbuster technology with a broad ability to mix and match genes; its use or misuse has profound implications for global ecology and the food supply. It is in no sense 'substantially equivalent' to plant breeding.

That sin may shortly be expunged. On 2 July, John Holdren, science adviser to US President Barack Obama, directed regulators to revisit the US framework for regulating agricultural biotechnology. Holdren is promising simpler rules for small producers, but also more transparency. Many US consumers have grown sceptical of the technology; in April 2014, Vermont became the first state to mandate labelling of products that contain GM crops. (The US House of Representatives has responded by passing a bill that would prohibit such state provisions.)

Some critics still hope that universal labelling on food packaging means the beginning of the end for GM crops. More probably, it will mark the end of the beginning — if it prises out a fresh approach from the scientific community and the agricultural biotechnology industry to come clean with the public on what they're doing. ■

Colin Macilwain writes about science policy from Edinburgh, UK.
e-mail: cfmworldview@googlemail.com

**SUBSTANTIAL
EQUIVALENCE
WAS THE
ORIGINAL SIN
THAT UNDERMINED
PUBLIC
CONFIDENCE
IN GM TECHNOLOGY.**

➔ **NATURE.COM**
Discuss this article
online at:
go.nature.com/n88nk3

RESEARCH HIGHLIGHTS

Selections from the
scientific literature

APPLIED PHYSICS

A low-power light amplifier

A semiconductor device can amplify the tiny signal from incoming photons using much less power and creating much less noise than current methods.

Previously, converting a signal from photons into a usable electrical signal required the use of two devices running at relatively high voltages — one device to convert the photon to an electrical signal and another to amplify it.

Yu-Hwa Lo at the University of California, San Diego, and his colleagues exploited a different amplification mechanism to consolidate the process into a single device. By engineering a special junction between layers of silicon with two different kinds of impurities, the team amplified the light signals by more than a factor of 4,000 and induced 30 times less noise in the signal than conventional methods.

Appl. Phys. Lett. 107, 053505 (2015)

AGRICULTURE

Pig-farming history traced

Domesticated pigs (pictured) routinely interbred with wild boars — contrary to common assumptions that humans kept their animals isolated.

Humans domesticated pigs from wild boars



independently in Anatolia (modern-day Turkey) and East Asia around 9,000 years ago. To learn about pig-population histories, a team led by Laurent Frantz at the University of Oxford, UK, analysed the genomes of more than 600 modern pigs and wild boars. After initial domestication in Anatolia, the ancestors of European pigs interbred with at least two different populations of wild boars that ranged between Europe and Anatolia. Pigs from East Asia seem to have interbred with local boars too. Despite this wild mixing, domestic pig genomes show signs of positive selection at regions that include genes involved in

behaviour and anatomy.

The researchers propose that ancient herders repeatedly selected pigs with useful traits, counteracting the effects of the wild boar genes.

Nature Genet. <http://dx.doi.org/10.1038/ng.3394> (2015)

ECOLOGY

Coral foe becomes a friend

Seaweed often inhibits the growth of corals, but it can help them when they are faced with a coral-eating starfish.

Seaweed can suppress coral growth by shading it from sunlight and by releasing toxic chemicals. Cody Clements and Mark Hay at the Georgia

over 35 years, Jo Clarke at the University of Glasgow, UK, and her colleagues found that the proportion of caught fish with no commercial value increased significantly below 600 metres.

Although this evidence supports a 600-metre depth limit in the northeast Atlantic, its relevance to other fishing areas is untested.

Curr. Biol. <http://doi.org/66n> (2015)

For more on this story, see go.nature.com/yjedl3



FISHERIES

Finding a limit for deep-sea fishing

The negative ecological impact of trawling for fish at depths of more than 600 metres outweighs the commercial benefits.

Deep-sea fish are particularly vulnerable to overfishing because populations grow slowly. This has led to calls for a maximum depth for trawling, but it has not been clear what that limit should be. By examining species from scientific surveys of the northeast Atlantic

Institute of Technology in Atlanta surrounded branches of a coral species (*Montipora hispida*) near Fiji with varying numbers of fronds of a common brown alga (*Sargassum polycystum*). After 4 months, they found that the growth rate of coral branches unencumbered by the seaweed was 2.7 times higher than corals with 8 surrounding fronds (the highest number tested).

However, coral branches surrounded by four or more seaweed fronds were rarely attacked by the crown-of-thorns starfish (*Acanthaster planci*), which devoured all the exposed corals.

Proc. R. Soc. B 282, 20150714 (2015)

CHRIS FURLONG/GETTY

BILL LILLING/GETTY

STEM CELLS

How stem cells tell signal from noise

Mouse stem cells switch on a neural developmental program when the activity of a specific gene lasts for a certain length of time.

Cells are flooded with many signals from gene expression. To find out how cells pick out the important ones from the background noise, Matt Thomson and his colleagues at the University of California, San Francisco, engineered mouse embryonic stem cells so that the *Brn2* gene turned on when it was exposed to light. When the activity of this gene reached a specific duration and level, the stem cells rapidly began specializing into neural progenitor cells.

Mathematical modelling showed that a positive feedback network in the *Brn2* circuitry helps to ensure that the *Brn2* signal rises above the noise.

Cell Systems 1, 117–129 (2015)

METEOROLOGY

Big coastal storms to come

Three major coastal cities on different continents could get walloped by tropical cyclones during the next century because of climate change.

Ning Lin of Princeton University in New Jersey and Kerry Emanuel of the Massachusetts Institute of Technology in Cambridge ran statistical models of how storms form near the cities of Dubai, Cairns in Australia and Tampa, Florida — all of which are vulnerable to rising sea levels. With climate change, storm surges could reach as high as 6 metres in Dubai during the next 100 years; the city has never experienced a tropical cyclone. Storm surges would be slightly lower in Cairns and Tampa, but still greater than the levels those cities have seen before.

The authors calculate that by the end of the century, the annual probability of such

powerful storms in Tampa would increase from about 1 in 10,000 now to between 1 in 2,500 and 1 in 700. *Nature Clim. Change* <http://dx.doi.org/10.1038/nclimate2777> (2015)

POPULATION BIOLOGY

Cane toads wage chemical war

Invasive toads in Australia could be turned against each other to control the population.

Richard Shine at the University of Sydney in Australia and his colleagues grew cane toad (*Rhinella marina*) tadpoles and embryos together in containers in the laboratory, and separated them with a mesh partition. They found that the tadpoles suppressed embryo growth by 33–84% and reduced their survival to less than 5%.

The tadpoles seem to produce a chemical that blocks embryo growth, allowing them to outcompete the embryos. The practice of removing tadpoles from breeding ponds could actually boost the growth of embryos; instead, tadpoles could be kept in mesh containers in the pond to stymie the embryos' growth, the authors say.

J. Appl. Ecol. <http://doi.org/635> (2015)

OBESITY

Genetic switch stores up fat

A variation in a genetic region associated with obesity causes fat to be stored rather than burned.

Melina Claussnitzer at the Beth Israel Deaconess Medical Center in Boston, Massachusetts, Manolis Kellis at the Massachusetts Institute of Technology in Cambridge and their colleagues studied fat cells from 52 people with a version of the *FTO* genetic region that is associated with obesity and from 48 people with the non-risk version. The team found a single-nucleotide change in the risk-associated *FTO* region that boosted

SOCIAL SELECTION

Popular topics on social media

Lifetime collaborators reap benefits

Few researchers can do science single-handedly, making collaborations crucial. According to an analysis in *Proceedings of the National Academy of Sciences*, long-term collaborations pay big dividends, yielding a 17% boost in citation rate for resulting papers. Using a metric based on publications between collaborators over time, the analysis identified a group of strong partnerships, or 'super ties', that produced an unusually high number of papers in a given period. Papers from authors with such ties receive an average of 21 more citations in biology and 8 more citations in physics than those without a super tie. The findings were well received by observers on social media. "Collaboration works. Pick your collaborators carefully, they can stay with you for a long time," tweeted Wouter Gerritsma, an information specialist at Vrije University in Amsterdam.

With those rewards in mind, Jeremy Borniger, a neuroscience

PhD student at Ohio State University in Columbus, used Twitter to ask: "Who wants to be my science life partner?" *Proc. Natl Acad. Sci. USA* 112, E4671–E4680 (2015)

➔ **NATURE.COM**
For more on popular papers:
go.nature.com/hkionq

expression of the genes *IRX3* and *IRX5*, which decreased the amount of energy burned and dissipated as heat. In fat precursor cells, this change resulted in the development of more energy-storing white fat cells and fewer energy-burning beige fat cells.

Inhibiting *Irx3* in mice caused the animals to lose weight without a change in appetite or exercise.

N. Engl. J. Med. <http://doi.org/6z5> (2015)

ECOLOGY

Apes get by in degraded habitat

Endangered chimpanzees could be adapting to landscapes that have been broken up by human activity.

Eastern chimpanzees (*Pan troglodytes schweinfurthii*) were thought to exist in low numbers in an area of Uganda where forest has been fragmented by farms and villages (pictured). Maureen McCarthy and Linda Vigilant of the Max Planck Institute for Evolutionary Anthropology in Leipzig, Germany, and their colleagues collected and

genetically analysed hundreds of chimpanzee droppings from about 630 square kilometres of fragmented habitat. They estimate that some 260 chimps in 9 communities live in about 1,200 square kilometres of what seems to be marginal habitat at best — more than 3 times the number of chimps that were previously estimated to live in this habitat.

The authors suggest that these and other rare species might be more adaptable — at least in the short term — than was thought.

BMC Ecol. <http://doi.org/66q> (2015)



➔ **NATURE.COM**
For the latest research published by Nature visit:
www.nature.com/latestresearch

SEVEN DAYS

The news in brief

EVENTS

Explosion arrests

The Chinese authorities have detained 12 people and are investigating another 11 in relation to a warehouse explosion in Tianjin earlier this month, state media report. Among those under scrutiny are senior executives of Rui Hai International Logistics, which owns the warehouse, the head of the Tianjin Municipal Transportation Commission, the president of Tianjin Port and a senior official with the Chinese transport ministry. As of 31 August, the disaster's death toll was 158. Questions have arisen over the handling and storage of chemicals implicated in the explosions.

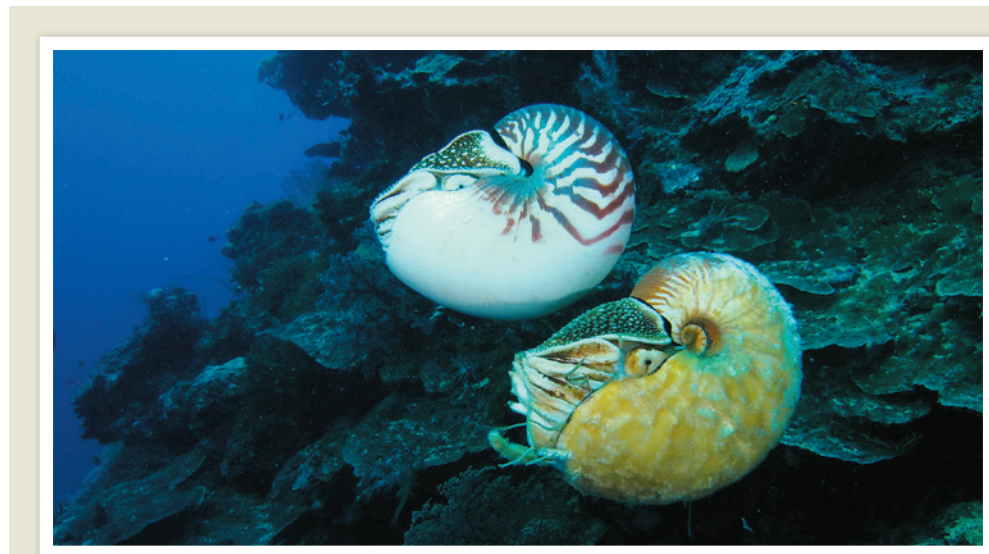
Iran-science uplift

Iran plans to boost international science collaboration once sanctions are lifted, Mohammad Farhadi, the country's science minister, said on 27 August. Speaking to Iran's Islamic Republic News Agency, Farhadi reported that preparations are under way to increase the country's cooperation with foreign universities, including the development of academic exchange programmes and a visit from an Austrian university delegation. Sanctions have hampered scientists' movement in and out of Iran, as well as the country's involvement in international projects, but will end once a nuclear-deal agreement made on 14 July is implemented.

FUNDING

Australian merger

The digital-research arm of Australia's Commonwealth Scientific and Industrial Research Organisation (CSIRO) will merge with National Information



PETER WARD

Slimy sea creature rears head once more

Thirty years after its last sighting, a rare species of nautilus (*Allonautilus scrobiculatus*) has been spotted near Papua New Guinea. A type of cephalopod, the species is distantly related to squid and octopuses, and was first discovered in the 1980s by Peter Ward of the University of

Washington in Seattle and Bruce Saunders from Bryn Mawr College, Pennsylvania. On 25 August, the University of Washington announced that Ward had again glimpsed the species, with its distinctive, 'slimy' shell covering (pictured), while visiting Papua New Guinea in July.

Communications Technology Australia (NICTA) to form a CSIRO digital-innovation team, called Data61.

Announced on 28 August, the merger follows the Australian government's decision to halt NICTA's funding after June 2016. Funds for Data61 will come from CSIRO's already-stretched budget, itself subjected in 2014 to cuts of Aus\$115 million (US\$82 million) over four years. The merger could result in the loss of as many as 200 jobs.

Grant cash rejected

The University of Florida in Gainesville announced on 27 August that a US\$25,000 grant from agriculture giant Monsanto, originally earmarked for a science outreach programme, will instead be given to the

campus food bank. The statement comes after Florida plant scientist Kevin Folta faced public threats over his acceptance of the money, which *Nature* first reported on 6 August (see *Nature* <http://doi.org/66p>; 2015). Monsanto refused to accept the university's offer to return the money. There is no suggestion of wrongdoing or scientific misconduct by Folta.

RESEARCH

Bear-brain clues

Knut, the celebrity polar bear hand-reared at the Berlin Zoological Garden, had an autoimmune brain disease, according to the latest investigation into his death. Knut drowned in 2011, aged 4, after suffering an epileptic fit and falling into a pool. An autopsy at the time blamed

an unspecified encephalitis — brain inflammation. An analysis published on 27 August found antibodies that signify anti-NMDA receptor encephalitis, in which the immune system harms nerve cells (H. Prüss *et al. Sci. Rep.* 5, 12805; 2015). The treatable disease was first reported in humans in 2007, but was unknown in animals. Knut's case suggests that autoimmune brain diseases may be more common in mammals than was thought.

NASA's icy choice

NASA has chosen New Horizons' next probable target: an icy body called 2014 MU69. The spacecraft will fly by the object in 2019, making it the mission's second destination after the historic encounter with Pluto in July. Kuiper belt object 2014 MU69 is about

SARA KRULWICH/NY TIMES/REDUX/EYEVINE

45 kilometres across, and New Horizons will fly within 12,000 kilometres of its surface. According to the 28 August announcement, the spacecraft will ignite its engines in a series of four burns, beginning in late October this year, to set itself on course for the fly-by. See go.nature.com/ojobwg for more.

E-waste woes

Only 35% of Europe's annual 9.5 million tonnes of electrical and electronic waste is legally disposed of, according to a study funded by the European Union and released on 30 August. The remaining 65% is either exported, recycled under non-compliant conditions, scavenged for the valuable elements in the waste or thrown away with the ordinary rubbish. Electronic waste contains toxic metals including mercury, cadmium and chromium, as well as valuable materials that can be reused, including gold, silver, palladium and rare-earth metals. The report offers detailed recommendations, with particular emphasis on educating consumers.

PEOPLE

Oliver Sacks dies

Neurologist and author Oliver Sacks (pictured) died at his home in Manhattan on 30 August, aged 82. Sacks



worked at several clinics in New York, and much of his writing revolved around the curious cases that he encountered. He found worldwide fame when his 1973 book *Awakenings* — which described how he roused encephalitis patients from a coma-like state with the Parkinson's disease drug L-dopa — was made into a film in 1990. Last month, he created the Oliver Sacks Foundation to promote understanding of the human brain through narrative non-fiction like his own.

Misconduct ruling

Surgeon Paolo Macchiarini, who is famous for implanting synthetic tracheas into humans, was cleared of scientific misconduct charges on 28 August by the Karolinska Institute in Stockholm, where Macchiarini is a visiting professor. The

university's vice-chancellor overruled a finding by an independent investigator that seven papers contained unsupported assertions of the success of artificial grafts. The decision to overrule was based on more than 1,000 pages of documents that Macchiarini and his co-authors submitted after the independent report was released. See go.nature.com/ynxom8 for more.

BUSINESS

Monsanto backs off

The world's largest seller of seeds, Monsanto, announced on 26 August that it has abandoned its US\$4.65-billion takeover attempt of Swiss competitor Syngenta after months of discussions. Executives of pesticide specialist Syngenta rejected multiple offers, despite pressure to negotiate a deal from some of their shareholders. The proposed merger was opposed by farmers' unions, who said that the move would reduce competition and push up seed and pesticide prices at a time when farmers' profits are under pressure from low food prices.

Sanofi Googled

The technology giant Google, based in Mountain View, California, announced on 31 August that it will collaborate with Paris-based

COMING UP

3–6 SEPTEMBER

Neurologists gather at the World Congress on NeuroTherapeutics: Dilemmas, Debates, Discussions, in Prague. go.nature.com/zmysnp

5–6 SEPTEMBER

Robotics engineers battle it out to find who has the best rover, at the European Rover Challenge near Chęciny, Poland. roverchallenge.eu/en

5–8 SEPTEMBER

Molecular biologists convene at EMBO's 6th meeting, in Birmingham, UK. the-embo-meeting.org

drugmaker Sanofi on ways to better monitor and treat people with diabetes. Sanofi is the second pharmaceutical company that Google has partnered with. In July 2014, it revealed a licensing deal with Novartis, based in Basel, Switzerland, for Google's 'smart lens' technology, which is designed to monitor glucose in tears. Financial terms of the Sanofi deal were not disclosed.

Gene-therapy trial

A first-of-its-kind clinical trial to test a treatment for a degenerative disease that causes blindness was given a green light by the US Food and Drug Administration on 24 August. The trial will test a treatment for retinitis pigmentosa devised by RetroSense Therapeutics of Ann Arbor, Michigan. The gene therapy will attempt to deliver a gene encoding a protein for light sensitivity called channelrhodopsin-2, which the firm hopes will make new light-sensing proteins in retinal cells. The trial will start by the end of the year.

NATURE.COM

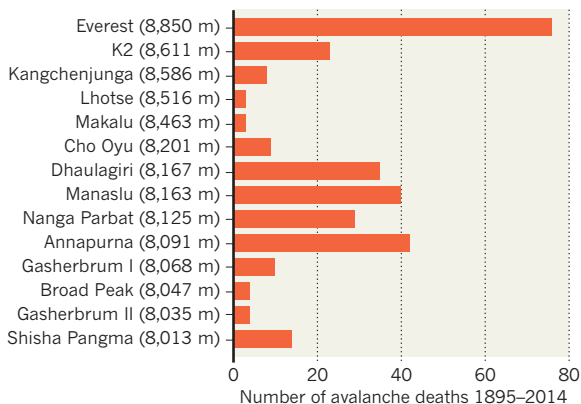
For daily news updates see: www.nature.com/news

TREND WATCH

Poor decision-making, improper camp placement and lack of good forecasting were to blame for 75% of avalanche fatalities on the world's highest mountains, says a report from the University of British Columbia in Vancouver, Canada (D. M. McClung *Ann. Glaciol.* <http://doi.org/66c>; 2015). Analysis of 10,000 mountaineering reports for Asian mountains during 1895–2014 found that 300 people died in avalanches while trying to climb peaks higher than 8,000 metres, one-third of the fatalities.

ASIAN MOUNTAINS' TERRIBLE TOLL

Camp placement in high-risk areas and lack of weather forecasting are behind 75% of avalanche deaths on peaks higher than 8,000 metres.

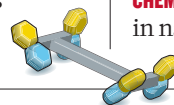


NEWS IN FOCUS

PHYSICS Entanglement can't be explained by local effects **p.14**

NORTHERN LIGHT Fourth-generation synchrotron powers up in Sweden **p.15**

ETHICS Panel considers data-sharing issues in big US study **p.16**



CHEMISTRY The building boom in nanoscale motors, rotors, switches and pumps **p.18**

NASA/JHUAPL/SWRI



The New Horizons craft photographed Pluto's atmosphere, backlit by the Sun, as the probe sailed away from the dwarf planet in mid-July.

PLANETARY SCIENCE

Pluto pressure data pose an atmospheric conundrum

Discrepancy arises between New Horizons and Earth-based measurements.

BY ALEXANDRA WITZE

NASA's New Horizons spacecraft solved many mysteries about Pluto when it flew past the dwarf planet in July. But as mission controllers prepare to steer the probe to its next rendezvous, planetary scientists are working to understand a puzzling result: an atmospheric pressure at Pluto's surface that is much lower than indicated by

measurements obtained from Earth.

Some have suggested that Pluto's atmospheric pressure is dropping as the dwarf planet's orbit carries it farther from the Sun and gases freeze out and fall to the surface as snow. But the most recent data taken from Earth suggest no such dramatic transformation. "I feel pretty secure that Pluto isn't starting to freeze out," says Eliot Young, a planetary scientist at the Southwest Research Institute

(SwRI) in Boulder, Colorado.

On 29 June, a few weeks before the fly-by, Young organized astronomers across New Zealand and Australia to watch Pluto as it passed in front of a distant star. Tracking how the star's light faded during the passage provided information on how much gas is in Pluto's atmosphere. Using the same method, planetary scientists have seen the atmosphere grow denser since 1988 — and analysis of ►

► the 29 June observations shows that the trend remains intact. Young calculates that the current atmospheric pressure at Pluto's surface is 22 microbars (0.022 pascals), or 22-millionths the pressure at sea level on Earth.

But on 14 July, New Horizons measured Pluto's surface pressure as much lower than that — just 5 microbars. “How we link the two, we're still working on,” says Cathy Olkin, a deputy project scientist for New Horizons at SwRI.

Part of the discrepancy between the spacecraft's observation and past estimates could be due to the indirect way that astronomers derive the value from Earth-based observations. These studies measure pressure some 50–75 kilometres above the dwarf planet's surface, and researchers use assumptions about the atmosphere's structure to calculate what that number translates to at the ground.

By contrast, New Horizons measured surface pressure directly by determining how strongly radio waves, beamed from antennas on Earth, bent as they passed through Pluto's atmosphere and arrived at the spacecraft on the far side of the dwarf planet.

“We may be looking at the first test of these models, not an atmospheric collapse.”

The next challenge is to figure out which of several competing models that describe Pluto's atmosphere can best reconcile the Earth-based measurements and what New Horizons measured at the surface.

“We may be looking at the first test of these models, not an atmospheric collapse or some spectacularly freaky physics,” says Ivan Linscott, a physicist at Stanford University in California and co-leader of the New Horizons radio measurement. “The jury's still out.”

Clues may yet come from New Horizons. About 95% of the data collected in its Pluto fly-by, including much of the information from the radio measurement, is still on board. Slow transmission speeds mean that the team will have to wait months for the rest of it to arrive. The transmission of images, which has been on pause since soon after the 14 July fly-by, will resume on 5 September.

And in late October, mission controllers will ignite the spacecraft's engines in a series of burns to set it on course for its next destination: an object called 2014 MU69, which is about 45 kilometres across and lies in the Kuiper belt, a collection of small bodies orbiting beyond Neptune. New Horizons is set to pass within about 12,000 kilometres of the object on New Year's Day 2019. ■

QUANTUM PHYSICS

Toughest test yet for quantum ‘spookiness’

Experiment plugs loopholes in previous demonstrations of ‘action at a distance’ and could make data encryption safer.

BY ZEEYA MERALI

It's a bad day both for Albert Einstein and for hackers. Physicists say that they have made the most rigorous demonstration yet of the quantum ‘spooky action at a distance’ effect that the German physicist famously hated — in which manipulating one object instantaneously seems to affect another one far away.

The experiment could be the final nail in the coffin for theories that are more intuitive than standard quantum mechanics. It could also enable engineers to develop a new suite of ultrasecure cryptographic devices. “From a fundamental point of view, this is truly history-making,” says Nicolas Gisin, a quantum physicist at the University of Geneva in Switzerland.

In quantum mechanics, objects can be in multiple states simultaneously: an atom can be in two places at once, for example. Measuring an object forces it to snap into a well-defined state. The properties of different objects also can become ‘entangled’, meaning that when one such object is measured, the state of its entangled twin also becomes set.

This idea galled Einstein because it seemed that this ghostly influence would travel instantaneously — contravening the universal rule that nothing can travel faster than the speed

of light. He proposed that quantum particles do have set properties, called hidden variables, before they are measured, and that even though those variables cannot be accessed they pre-program entangled particles to behave in correlated ways.

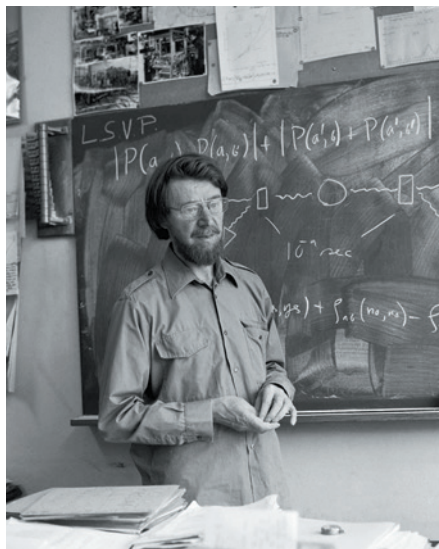
In the 1960s, physicist John Bell proposed a test that could discriminate between Einstein's hidden variables and spooky action at a distance¹. He calculated that hidden variables can explain correlations only up to some maximum limit. If that level is exceeded, then Einstein's model must be wrong.

The first experiment suggesting that this was the case was carried out in 1981 (ref. 2). Many more have been performed since, always coming down on the side of spookiness — but each has had loopholes that meant that physicists have never been able to fully close the door on Einstein's view. Experiments that use entangled photons are prone to the ‘detection loophole’: not all photons produced in the experiment are detected, and sometimes as many as 80% are lost. Experimenters therefore have to assume that the photons they capture are representative of the entire set.

To get around the detection loophole, physicists often use particles that are easier to keep track of than are photons, such as atoms. But it is tough to place atoms far apart without destroying their entanglement. This opens the ‘communication loophole’: if the entangled atoms are too close together, then, in principle, measurements made on one could affect the other without violating the speed-of-light limit.

ENTANGLEMENT SWAPPING

In the latest paper³, which was submitted to the arXiv preprint repository on 24 August and has not yet been peer reviewed, Ronald Hanson of Delft University of Technology and his colleagues report the first Bell experiment that closes both the detection and the communication loopholes. The team used a cunning technique called entanglement swapping to combine the benefits of using both light and matter. The researchers started with two unentangled electrons sitting in diamond crystals in different labs on the Delft campus, 1.3 kilometres apart. Each electron was individually entangled with a photon,



John Bell devised a test to show that nature does not ‘hide variables’ as Einstein had proposed.

and both those photons were then zipped to a third location. There, the two photons were entangled with each other — and this caused both their partner electrons to become entangled, too.

This did not work every time. In total, the team managed to generate 245 entangled pairs of electrons over the course of nine days. The team's measurements exceeded Bell's bound, once again supporting the standard quantum view. Moreover, the experiment closed both loopholes at once: because the electrons were easy to monitor, the detection loophole was not an issue, and they were separated by enough distance to also close the communication loophole.

"It is a truly ingenious and beautiful experiment," says Anton Zeilinger, a physicist at the Vienna Centre for Quantum Science and Technology.

Matthew Leifer, a quantum physicist at the Perimeter Institute for Theoretical Physics in Waterloo, Canada, says that he would not be surprised to see one of the authors of the paper share a Nobel prize in the next few years. "It's that exciting."

A loophole-free Bell test also has implications for quantum cryptography, says Leifer. Companies already sell systems that use quantum mechanics to block eavesdroppers. The systems produce entangled pairs of photons, sending one photon in each pair to one user and the other photon to a second user. The two users then turn these photons into a cryptographic key that only they know.

But loopholes — and the detection loophole in particular — mean that malicious companies could sell devices that fool users into thinking that they are getting quantum-entangled particles, when they are instead being given keys that the company can use to spy on them. In 1991, quantum physicist Artur Ekert observed⁴ that integrating a Bell test into the system would ensure a genuine quantum process. For this to be valid, however, the Bell test must be free of any loopholes. The Delft experiment "is the final proof that quantum cryptography can be unconditionally secure", says Zeilinger.

In practice, the technique will be hard to implement, because so far it has generated entangled electrons at a very slow pace.

Zeilinger also notes that there remains a last, somewhat philosophical, loophole, first identified by Bell himself: the possibility that hidden variables could somehow manipulate the experimenters' choices of what properties to measure, tricking them into thinking quantum theory is correct. ■

1. Bell, J. S. *Physics* **1**, 195–200 (1964).
2. Aspect, A., Grangier, P. & Roger, G. *Phys. Rev. Lett.* **49**, 91–94 (1982).
3. Hensen, B. *et al.* Preprint available at <http://arxiv.org/abs/1508.05949> (2015).
4. Ekert, A. *Phys. Rev. Lett.* **67**, 661–663 (1991).



The next-generation synchrotron at Lund in Sweden has passed its first test.

TECHNOLOGY

X-ray science gets an upgrade

Swedish synchrotron promises super-bright beams and will open up new avenues for researchers.

BY DAVIDE CASTELVECCHI

In what researchers hope marks the start of a new era for X-ray science, electrons have begun circulating in a next-generation synchrotron in Lund, Sweden. This machine promises to lower the costs of X-ray-light sources around the world, while improving their performance and enabling experiments that were not possible before.

Synchrotrons are particle accelerators that produce X-rays that are used in research ranging from structural biology to materials science. At 10 p.m. local time on 25 August, the first bunches of electrons began circulating inside a new 528-metre-long, 3-gigaelectron-volt (GeV) machine at the MAX IV facility in Lund, project director Christoph Quitmann told *Nature*. MAX IV is the first 'fourth-generation' synchrotron in the world.

Getting the first beam is an absolutely crucial first step" in demonstrating fourth-generation technology, says Chris Jacobsen, an X-ray physicist at the Argonne National Laboratory in Lemont, Illinois. MAX IV, he says, is "leading the world towards a new path in synchrotron light sources".

In synchrotrons, bunches of electrons circulate at nearly the speed of light inside a ring-shaped vacuum tube. Powerful 'bending' magnets steer the electrons around the rings, and 'focusing' magnets push them together against their mutual repulsion. The electrons then pass through special magnets that shake them sideways to produce pulses of X-rays, known as synchrotron radiation.

Fourth-generation light sources promise to squeeze the electrons into tighter bunches, leading to X-ray pulses that concentrate more photons into a tighter, brighter beam. This means that it will take just minutes for researchers to do experiments that could take days on a third-generation machine, Jacobsen says.

FOURTH GENERATION

Eventually, beams from fourth-generation machines could enable materials scientists to observe chemical reactions inside a battery as they happen, or structural biologists to reveal the structure of proteins from smaller protein crystals than those needed at existing light sources.

The crucial innovation in the fourth-generation machines is to employ a narrower ►

► vacuum pipe in which to circulate the electrons. In MAX IV's case, the pipe is 22 millimetres across, about half as wide as in a typical existing synchrotron. This makes it possible to get stronger magnetic fields using more-compact bending and focusing magnets, which are also less expensive and can consume ten times less electricity than third-generation systems because of their smaller size.

But keeping such a narrow pipe free of air would not have been possible using conventional high-vacuum pumps alone. MAX IV borrowed a technology from the Large Hadron Collider (LHC) at CERN, Europe's particle-physics facility near Geneva, Switzerland, which circulates protons rather than electrons. The LHC's trick — now adopted by MAX IV — is to coat the inner surface of the pipes with a special alloy that absorbs any gas molecules that happen to bounce around inside the tubes.

"The Swedes should be very proud of their innovative fabrication techniques, which lower the cost of making these machines," says physicist Herman Winick, a veteran synchrotron builder at the SLAC National Accelerator Laboratory in Menlo Park, California.

In the next few weeks, the MAX IV team will have to test whether they can circulate the large number of electrons that will be necessary to produce high-quality beams of X-rays, says Robert Hettel, an accelerator physicist at SLAC. And in subsequent months, they will build eight experimental stations, or beamlines, around the synchrotron, which they plan to open on 21 June 2016, a date chosen for the symbolism of the summer solstice.

The synchrotron that fired up on 25 August is the larger of two that the MAX IV team is building; the smaller fourth-generation machine will produce electrons of 1.5 GeV for making 'softer', or less energetic, X-rays. The combined cost of the machines and of the first eight beamlines will be 4.5 billion Swedish kronor (US\$530 million), Quitmann says, which is being paid for by the Swedish government.

Quitmann says that his team reached "a major milestone last night". But, he adds, "We have still a long way to go". ■



The US Precision Medicine Initiative aims to collect health and genetic data from 1 million people.

PERSONALIZED MEDICINE

Health study set to decide data policy

Specialists are split over whether participants should have free access to their genetic information.

BY SARA REARDON

After dozens of unsuccessful treatments, Eric Dishman started to suspect that his illness was due to something other than the rare kidney cancer he was diagnosed with in 1989. Five years ago, he had his whole genome sequenced, then gave the data to oncologists — and learned that he had a different type of cancer altogether.

He was treated successfully, and remains cancer-free. "I was an early prototype for precision medicine," he says.

Dishman now leads the health and life-sciences division of microprocessor giant Intel in Banks, Oregon. He is also a member of a working group run by the US National Institutes of Health (NIH) for the Precision Medicine Initiative (PMI) — a US\$215-million project to collect data on genomes, health records and physiological measurements from

1 million participants, to learn how genetics, environment and lifestyle influence disease risk and the effectiveness of treatments.

Next month, the group is expected to release a project plan. Observers are eager to learn its answer to a key question: how much information about disease risk, especially genetic data, will the project share with participants?

That issue is the subject of much debate. Dishman and others say that participants should at least have the option to see all their personal data so that they can investigate their own health, just as he did. But some specialists in the field say that showing participants their data is irresponsible, because the information is challenging for people to interpret and its significance is often uncertain.

Most genetic variants linked to disease increase risk only slightly, yet people who discover that their genome holds such a variant might worry excessively or seek unnecessary

ANDY KROPA/REDUX/EYEVINE



**MORE
ONLINE**

TOP STORY



Pioneer behind controversial PubPeer site reveals his identity go.nature.com/t5edtq

MORE NEWS

- Over half of psychology studies fail reproducibility test go.nature.com/iz5eqq
- Evidence supports limit on deep-sea trawling go.nature.com/yjed13
- Space station dark-matter experiment hits a glitch go.nature.com/v8emet

NATURE PODCAST



Rethinking the water cycle; weather forecasting; and the Podcast's 400th episode nature.com/nature/podcast

medical tests. Or they might do nothing: the limited research on how people react suggests that, far from causing panic, information about common variants of small-to-moderate effect does not seem to motivate people to make recommended long-term behavioural changes to lessen risk. “Unless you give people the tools and the skills to deal with the raw data, I don’t see how you could give them the raw data,” says Brian Van Ness, a geneticist at the University of Minnesota in Minneapolis.

Others counter that letting researchers choose what information to share with participants is paternalistic. “I don’t know why we’re so afraid of the genome that we say it’s dangerous for people to have this information,” says Sharon Terry, director of the Genetic Alliance in Washington DC. “All of us live with all kinds of uncertainty all the time.”

The NIH, together with US President Barack Obama, who officially launched the PMI in his January 2015 State of the Union speech, says that it wants participants to be active partners in the research. And given the programme’s large size and high profile, the working group’s choices could establish a precedent for similar projects. “What’s implemented in this cohort will be the future of research,” says NIH working-group co-chair Bray Patrick-Lake, who works in patient engagement in research at Duke University in Durham, North Carolina.

Dishman, for his part, favours a tiered system in which each participant decides how much data to receive. Although he credits his eventual cancer cure to his analysis of his own genome, he says that not everyone has the education or the desire to interpret genetic information. If the PMI achieves its goal of recruiting a diverse population, Dishman says that it will probably include some participants who would rather not receive their genetic data. “There are so many psychological, emotional and cultural responses, and you can’t just create blanket policies,” he says.

Patrick-Lake says that the group will probably set a single policy for the whole initiative, although discussing what that will be would be premature.

If the working group does decide to give participants access to their data, it will have to address many issues related to the transaction. US law requires genetic data used for clinical decision-making to be of higher quality — and thus more costly to produce — than data used only for research. So sharing data could drive up costs. The project will also need to decide on questions such as whether to release data to the families of participants who have died.

Historically, most studies have not given

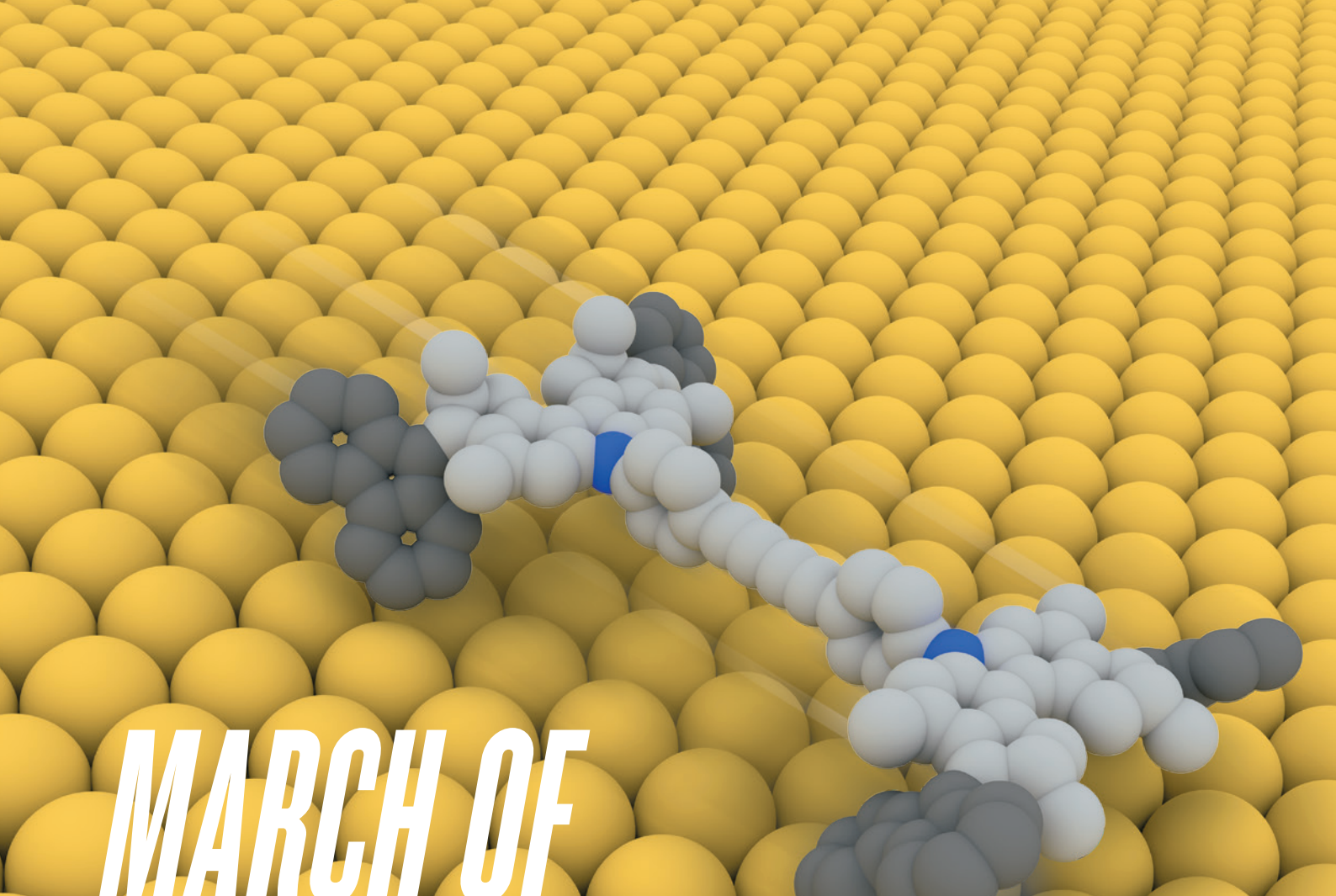
“All of us live with all kinds of uncertainty all the time.”

raw data to individuals, especially if the findings do not reveal any information that participants could use to improve their health. But public opinion on data sharing seems to be shifting towards openness, says Susan Wolf, who studies public-health law and bioethics at the University of Minnesota. “We’re past the era when scientists can simply take specimens, generate data with great health importance, and decline to offer any of that data back to people.”

The UK Biobank in Stockport, which is enrolling 500,000 people, collects genomic and health data but does not return individual results to participants. That was the simplest logical approach when the biobank was set up in 2007, says executive director Rory Collins, an epidemiologist at the University of Oxford, UK, who is a member of the NIH working group. “If the UK Biobank was set up now, it’s plausible that a different decision would be taken,” he says. “I think the Precision Medicine Initiative may take a different decision.” ■ [SEE EDITORIAL P.5](#)

CORRECTION

The News Feature ‘The cannabis experiment’ (*Nature* **524**, 280–283; 2015) incorrectly located the University of Colorado School of Medicine in Denver. It is actually in Aurora.



MARCH OF THE MACHINES

BY MARK PELOW

INSPIRED BY BIOLOGY, CHEMISTS HAVE CREATED A CORNUCOPIA OF MOLECULAR PARTS THAT ACT AS SWITCHES, MOTORS AND RATCHETS. NOW IT IS TIME TO DO SOMETHING USEFUL WITH THEM.

The robot moves slowly along its track, pausing regularly to reach out an arm that carefully scoops up a component. The arm connects the component to an elaborate construction on the robot's back. Then the robot moves forward and repeats the process — systematically stringing the parts together according to a precise design.

It might be a scene from a high-tech factory — except that this assembly line is just a few nanometres long. The components are amino acids, the product is a small peptide and the robot, created by chemist David Leigh at the University of Manchester, UK, is one of the most complex molecular-scale machines ever devised.

It is not alone. Leigh is part of a growing band of molecular architects who have been inspired to emulate the machine-like biological molecules found in living cells — kinesin proteins that stride along the cell's microscopic scaffolding, or the ribosome that constructs proteins by reading genetic code. Over the past 25 years, these researchers have devised an impressive array of switches, ratchets, motors, rods, rings, propellers and more — molecular mechanisms that can be plugged together as if they were nanoscale Lego pieces. And progress is accelerating, thanks to improved analytical-chemistry tools and reactions that make it easier to build big organic molecules.

Now the field has reached a turning point. "We've made 50 or 60 different motors," says Ben Feringa, a chemist at the University of Groningen in the Netherlands. "I'm less interested in making another motor than actually using it."

That message was heard clearly in June, when one of the influential US Gordon conferences focused for the first time on molecular machines and their potential applications, a clear sign that the field

A molecular 'nanocar' travels across a metal surface, propelled by bonding changes.

KARL-HEINZ ERNST

has come of age, says the meeting's organizer, chemist Rafal Klajn of the Weizmann Institute of Science in Rehovot, Israel. "In 15 years' time," says Leigh, "I think they will be seen as a core part of chemistry and materials design."

Getting there will not be easy. Researchers must learn how to make billions of molecular machines work in concert to produce measurable macroscopic effects such as changing the shape of a material so that it acts as an artificial muscle. They must also make the machines easier to control, and ensure that they can carry out countless operations without breaking.

That is why many in the field do not expect the first applications to involve elaborate constructs. Instead, they predict that the basic components of molecular machines will be used in diverse areas of science: as light-activated switches that can release targeted drugs, for example, or as smart materials that can store energy or expand and contract in response to light. That means that molecular architects need to reach out to researchers who work in fields that might benefit from their machine parts, says Klajn. "We need to convince them that these molecules are really exciting."

SHUTTLE LAUNCH

Many of today's molecular machines trace their origins to a relatively simple device built in 1991 by Fraser Stoddart, a chemist now at Northwestern University in Evanston, Illinois. It was an arrangement known as a rotaxane, in which a ring-shaped molecule is threaded onto an 'axle', a linear molecule capped by bulky stoppers at each end. Included in this particular axle, towards either end of the chain, were two chemical groups that could bind to the ring. Stoddart found¹ that the ring could hop back and forth between these two sites, creating the first molecular shuttle.

By 1994, Stoddart had modified the design so that the axle had two different binding sites². The shuttle existed in solution; changing the acidity of this liquid forced the ring to hop from one site to the other, making the shuttle into a reversible switch. Similar molecular switches could one day be used in sensors that respond to heat, light or specific chemicals, or that open the hatch of a nanoscale container to deliver a cargo of drug molecules at precisely the right time and to exactly the correct place in a person's body.

Stoddart's switches displayed two properties that would come up again and again in the molecular machines that followed. First, the links between the ring and the axle's binding sites were not the strong covalent bonds that knit atoms into molecules. Instead, they were weaker electrostatic attractions between slightly positive and negative regions of the two components. This meant that the bonds could be readily formed and broken, much like zipping and unzipping the hydrogen bonds that link the two strands of DNA. Second, the shuttles did

not need an external energy source to zip back and forth. They were powered by collisions with other molecules in the solution, a jostling effect called Brownian motion.

A plethora of other switches soon followed. Some were controlled with light or changes in temperature, whereas others worked by binding specific ions or molecules from solution, in a similar way to how ion channels work in cell membranes, opening or closing in response to chemical signals.

Stoddart, however, took his research in a

"I THINK THEY WILL BE SEEN AS A CORE PART OF CHEMISTRY AND MATERIALS DESIGN."

different direction. Working with James Heath at the California Institute of Technology in Pasadena, he used millions of rotaxanes to make a memory device³. Sandwiched between silicon and titanium electrodes, the rotaxanes could be electrically switched from one state to another and used to record data. This molecular abacus, roughly 13 micrometres across, contained 160,000 bits, each composed of a few hundred rotaxanes — a density of roughly 100 gigabits per square centimetre, comparable to the best commercial hard drives available today.

Using 24 of the best-performing bits, Stoddart's team stored and retrieved the letters 'CIT' (for the California Institute of Technology). But the switches were not very robust, typically falling apart after fewer than 100 cycles. One promising solution is to load them into tough, porous crystals known as metal-organic frameworks (MOFs), which protect the switches and organize them into a precise 3D array (see *Nature* 520, 148–150; 2015).

Earlier this year, Robert Schurko and Stephen Loeb of the University of Windsor, Canada, showed that they could pack about 10²¹ molecular shuttles into each cubic centimetre of a MOF⁴. And last month, Stoddart unveiled⁵ a different MOF that contained switchable rotaxanes. The MOF was mounted on an electrode, and the rotaxanes could be switched en masse by changing the voltage.

Researchers working on these MOFs hope that the 3D, solid scaffolds will offer a greater density of switches than conventional silicon transistors, and make the molecules easier to switch in a controllable way, potentially offering vast amounts of data storage. "The sci-fi way to think about it would be to address each molecule as a bit," says Loeb. But more realistically, he says, a speck of the MOF containing hundreds of switches could act as one bit. As long as most of the switches in the speck

function properly, he says, they will collectively and reliably encode data.

Others have used rotaxanes to make switchable catalysts. In 2012, Leigh described⁶ a system with a nitrogen atom in the middle of the rotaxane's axle, where it is normally covered by a ring. Add an acid, and the ring moves to one side, exposing the nitrogen atom so that it can catalyse a common chemical reaction. It goes further: last November, Leigh reported⁷ a rotaxane system with two different catalytic sites. Moving the ring from one to the other

allowed the chemists to switch the rotaxane's activity, so that it could stitch together a mixture of molecules in two different ways. Leigh is now working on putting several different switchable catalysts into the same solution, where they could be toggled on and off in a sequence to build target molecules into complex products, in much the same way as enzymes do in a cell.

NANO MOTORS

In 1999, after early experiments with shuttles and switches, the field took a big step forward with the creation of the first synthetic molecular motor⁸. Built by Feringa's team, it was a single molecule containing two identical 'paddle' units connected by a carbon-carbon double bond. This fixed the paddles in place until a burst of light broke part of the bond, allowing the paddles to rotate. Crucially, the shape of the paddles meant that they could turn in only one direction — and as long as there was a supply of light and some heat, the motor would just keep spinning.

Feringa went on to use similar molecular motors to create a four-wheel-drive 'nanocar'⁹. He also showed¹⁰ that the motors could give liquid crystals enough of a twist to slowly rotate a glass rod sitting on top of them. The rod was 28 micrometres long — thousands of times the size of the motors.

Some chemists argue that although these motors are cute, they are ultimately useless by themselves. "I've always been a bit sceptical of artificial motors — they're too difficult to make, too difficult to scale up," says Dirk Trauner, a chemist at the Ludwig Maximilian University in Munich, Germany.

But the chemical principles behind them might be very useful indeed. Using the same light-activated mechanism, researchers have developed around 100 drug-like compounds that can be switched on or off in response to light.

In July, for example, a team led by Trauner

reported¹¹ a light-switchable version of combretastatin A-4, a potent anticancer compound that comes with some serious side effects, because it indiscriminately attacks tumour cells and healthy tissue alike. The team's switchable drug could drastically reduce system-wide side effects: it contains a nitrogen–nitrogen double bond that holds two sections of the molecule apart and renders it inactive. Only under blue light will the bond break and allow the sections to rotate into the molecule's active form. Trauner says that an area of tissue just 10 micrometres wide can be specifically targeted in this way, using light delivered through a flexible tube or by an implanted device. Trauner is planning mouse studies to test the compound's effectiveness against cancer.

He also hopes to use photoswitchable compounds to restore vision in people with macular degeneration or retinitis pigmentosa, conditions that damage the eye's light-sensing rod and cone cells. "It's low-hanging fruit — because it's in the eye, you don't have to worry about how to get the light in," he says. Last year, he showed¹² that one injection of a photoswitchable molecule called DENAQ into the eyes of blind mice partially restored their vision for several days, allowing the animals to distinguish between light and dark. The team is now trying the same technique in primates, and hopes to begin human trials in two years' time.

Trauner and Klajn both agree that the main challenge will be to convince the cautious pharmaceutical industry that photoswitchable drugs have potential, even though they have no track record in humans. "We need to get the pharmaceutical industry excited about photopharmacy," says Trauner. "Once they see the value, we'll be in good shape."

WALK THE LINE

Long before any creature had evolved to move on dry land, cells were using legs as part of their cellular machinery. Prime examples are the two-pronged proteins called kinesins, which put one 'foot' in front of the other as they carry molecular cargo along the cell's stiff scaffolding of microtubules.

Inspired by kinesin, researchers have built artificial walkers from DNA. The molecules typically have feet that are anchored in place by binding to complementary strands of DNA laid out on a track; adding a competing DNA strand can free the foot, allowing it to take a step forward. One of the most striking examples was described¹³ in 2010 by Nadrian Seeman at New York University. His DNA walker had four 'feet' and three 'hands', with which it could pick up gold nanoparticles as it moved around a tile made of folded DNA.

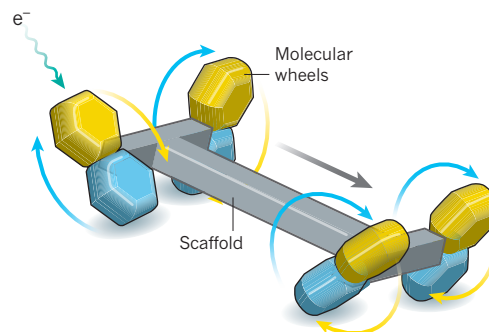
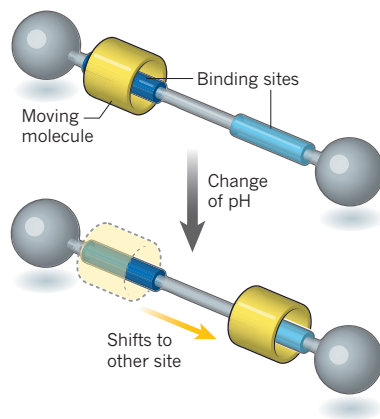
DNA walkers — and variants that soon trundled out of other labs — would wander aimlessly if they did not have a built-in ratchet system to stop them taking a step backward. For many walkers, that ratchet lies in the

NANO MACHINES

Mechanisms the size of molecules — governed by the rules of chemistry, rather than Newtonian mechanics — could have applications ranging from drug delivery to nanoscale computer memories.

MOLECULAR SWITCH

A ring-shaped molecule threaded onto a linear molecule shifts between two binding sites depending on the acidity of the surrounding solution.



NANOCAR

Electrons from a scanning tunnelling microscope tip (not shown) leap onto the molecules that form the 'wheels' of this device, causing them to change configuration, rotate and move the car forward.

relative rates of the chemical reactions that are involved in binding and releasing their feet, with the pummelling of Brownian motion driving the released foot forward¹⁴.

Over the past few years, detailed chemical studies and molecular dynamics simulations have shown that this 'Brownian ratchet' concept underlies all chemically driven molecular machines, including many biological motors. In 2013, for example, a team led by Nils Walter, a chemical biologist at the University of Michigan in Ann Arbor, found¹⁵ the same mechanism at work in the spliceosome, a cellular machine that snips sections out of RNA before genetic information is translated to make proteins. "Kinesin uses it, the ribosome uses it and the spliceosome uses it," says Walter.

That shows that the same principles underlie biological machines and synthetic molecular machines, so researchers working in the two areas could share knowledge. "By and large, they're quite separate fields right now," says Walter. "I think the next breakthroughs will come if we all sit at the same table."

ROCKET SCIENCE

Meanwhile, inspired by the microscopic medical submarine of the cult 1966 film *Fantastic Voyage*, chemists have created an array of micrometre-sized particles and tubes that can zip through liquids like rockets.

Some of these motors carry a catalyst that generates thrust by producing a stream of bubbles from the liquid around them — often hydrogen peroxide. Others get their power directly from light or from external electric

and magnetic fields, which can also be used to steer the vessels. "These nanomotors can go over 1,000 times their own length per second, it's incredible," says nanoengineer Joseph Wang of the University of California, San Diego. He thinks that the most promising applications lie in fast drug delivery, or low-cost clean-up of environmental pollutants — although many in the field caution that it is too early to tell whether nanomotors would trump conventional methods.

Hydrogen peroxide, a powerful oxidizing agent, is hardly conducive to *in vivo* use. "When all the work was based on peroxide there was a lot of scepticism," Wang admits. But in December last year, he reported¹⁶ a microscale motor suitable for testing in live animals. Made of a plastic tube roughly 20 micrometres long, it contains a core of zinc that reacts with stomach acid to generate propulsive bubbles of hydrogen.

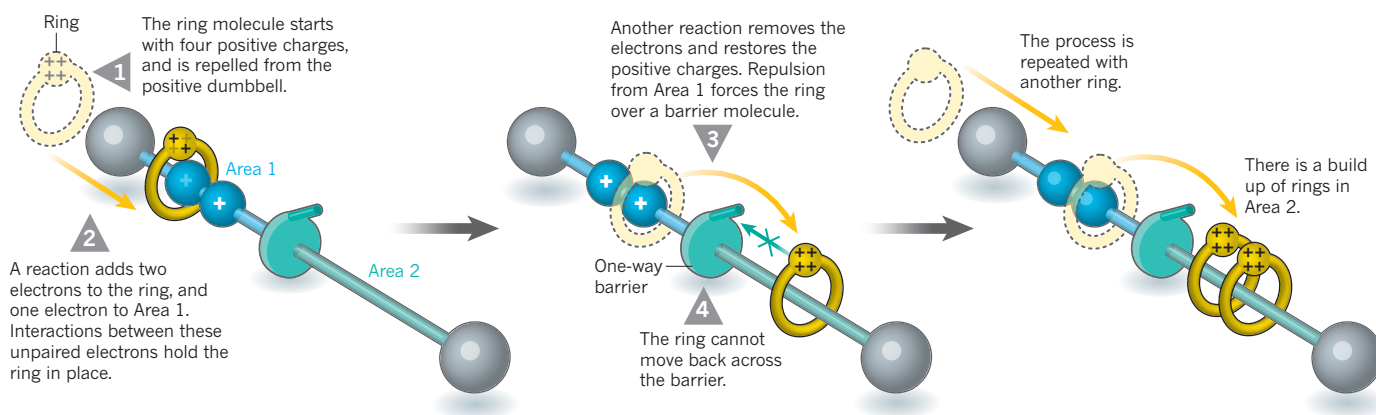
The tubes safely zipped around inside a mouse's stomach for about 10 minutes. Wang used them to carry gold nanoparticles into surrounding stomach tissue; mice dosed with plain nanoparticles ended up with three times less gold in their stomach lining than mice dosed with the tubes.

Wang suggests that loading drugs or imaging compounds onto the rockets could deliver them into stomach tissue rapidly and effectively. "In the next five years we will move to practical *in vivo* applications," he says. "It really is the fantastic voyage."

At the moment, there is limited crossover between research on these rockets and the molecular machines. "But we could bring

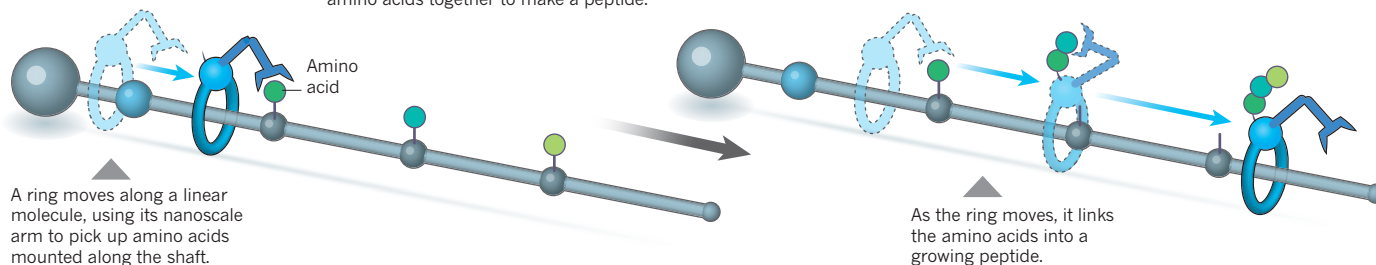
MOLECULAR PUMP

Chemical reactions in a surrounding solution drive two molecular rings into a holding area.



MOLECULAR ASSEMBLY LINE

A machine with a nanoscale 'arm' can connect amino acids together to make a peptide.



a lot," says Klajn. For example, coating a micromotor with light-responsive molecular switches could offer extra control over its movement, he suggests.

PUMP IT UP

In their quest to forge molecular machines that can actually do something useful, researchers are starting to integrate several different components into a single device. In May this year, Stoddart unveiled¹⁷ an artificial molecular pump that pulls two ring molecules out of solution onto a storage chain. Each ring slips over a stopper at one end of the chain, attracted to a switchable binding point. Flipping that switch pushes the ring over a second barrier farther along the chain, where it reaches a holding area (see 'Nano machines').

The system is not able to pump any other type of molecule, and it took a lot of trial and error to build. "It's been a long road," sighs Stoddart. But it proves that molecular machines can be used to concentrate molecules, pushing a chemical system out of equilibrium in the same way that biology can build up a store of potential energy by forcing ions or molecules up a concentration gradient. "We're learning how to design an energy ratchet," he says.

Stoddart says that such developments could enable the field to progress in two major directions: stay nano, giving the machines molecular-scale jobs that cannot be achieved in any other way; or go macro, using trillions of them together to reshape materials or move substantial cargoes, like an army of ants.

Perhaps the prime example of the nano

approach is Leigh's molecular assembly line¹⁸. Inspired by the ribosome, it is based on a rotaxane system that picks up amino acids from its axle and adds them to a growing peptide chain. But the devices could have macro applications. Over 36 hours, 10¹⁸ of them working together can produce a few milligrams of peptide. "It doesn't do anything that you can't do in the lab in half an hour," says Leigh. "Yet it shows that you can have a machine that moves down a track and picks up molecular building blocks and puts them together." Leigh is now working on other versions of the machine to make sequenced polymers, with tailored material properties.

Conversely, trillions of molecular machines working together could change the properties of materials in the macroscopic world. Gels that expand or contract in response to light or chemicals, for example, could act as adjustable lenses or sensors. "In the next five years, I bet you'll get the first smart materials where you have switches incorporated," says Feringa.

Rotaxane-like molecules are already starting to see commercial applications. The Nissan Scratch Shield iPhone case, launched in 2012 and based on work by Kohzo Ito at the University of Tokyo, is made of polymer strands threaded through pairs of barrel-shaped cyclodextrin molecules connected in a figure-of-eight shape. Pressure on a normal polymer coating would break the connections between the chains, leaving a scratch. But the cyclodextrin rings act like the wheels of a pulley system, allowing the polymer strands to slip through without breaking¹⁹. The films can

even protect a brittle screen from a sustained beating with a hammer.

For Stoddart, this shows that the components developed by molecular architects are already ripe for application. "This field has come a long way," says Stoddart. "Now we have to start showing it's useful." ■

Mark Peplow is a science journalist based in Cambridge, UK.

1. Anelli, P. L., Spencer, N. & Stoddart, J. F. *J. Am. Chem. Soc.* **113**, 5131–5133 (1991).
2. Bissell, R. A., Córdova, E., Kaifer, A. E. & Stoddart, J. F. *Nature* **369**, 133–137 (1994).
3. Green, J. E. *et al. Nature* **445**, 414–417 (2007).
4. Zhu, K., O'Keefe, C. A., Vukotic, V. N., Schurko, R. W. & Loeb, S. J. *Nature Chem.* **7**, 514–519 (2015).
5. McGonigal, P. R. *et al. Proc. Natl Acad. Sci. USA* <http://dx.doi.org/10.1073/pnas.1514485112> (2015).
6. Blanco, V., Carlone, A., Hänni, K. D., Leigh, D. A. & Lewandowski, B. *Angew. Chem. Int. Edn* **51**, 5166–5169 (2012).
7. Beswick, J. *et al. Chem. Sci.* **6**, 140 (2015).
8. Koumura, N., Zijlstra, R. W. J., van Delden, R. A., Harada, N. & Feringa, B. L. *Nature* **401**, 152–155 (1999).
9. Kudernac, T. *et al. Nature* **479**, 208–211 (2011).
10. Eelkema, R. *et al. Nature* **440**, 163 (2006).
11. Borowiak, M. *et al. Cell* **162**, 403–411 (2015).
12. Tochitsky, I. *et al. Neuron* **81**, 800–813 (2014).
13. Gu, H., Chao, J., Xiao, S.-J. & Seeman, N. C. *Nature* **465**, 202–205 (2010).
14. Astumian, R. D. *Nature Nanotechnol.* **7**, 684–688 (2012).
15. Krishnan, R. *et al. Nature Struct. Mol. Biol.* **20**, 1450–1459 (2013).
16. Gao, W. *et al. ACS Nano* **9**, 117–123 (2015).
17. Cheng, C. *et al. Nature Nanotechnol.* **10**, 547–553 (2015).
18. Lewandowski, B. *et al. Science* **339**, 189–193 (2013).
19. Noda, Y., Hayashi, Y. & Ito, K. *J. Appl. Polymer Sci.* **131**, 40509 (2014).

THE TROUBLE WITH WEARABLES

Electronic gadgets on — and in — our bodies are multiplying fast, but transmitting all their data safely will be a challenge.

BY KAT AUSTEN

Tom is late for his train and doesn't know the way to the station. Racing around a corner, he runs into a plaza full of tourists snapping and uploading photos to Instagram and Facebook. Which way should he go? He tells his Internet-connected contact lenses to load a map, meanwhile tapping at his smartwatch to pull up his ticket and platform information. An alarm flashes in his peripheral vision, only 15 minutes until the train departs, but the map is not loading. He looks around in dismay, frantically yelling "refresh" to his lenses against the clamour of the street. An alert scrolls across his vision: "You're feeling stressed. Take a breath. Have a hug!" But with all the tourists accessing the Internet, Tom has no hope of getting his much-needed map.

Welcome to the chaotic future of wearable electronics: devices that promise to connect real to digital lives seamlessly. These gadgets are rapidly multiplying, and within five years there could be half a billion devices strapped onto, or even embedded in, human bodies. Today, the most familiar gadgets are fitness trackers and smart watches, which monitor health and provide ready access to online services. But there are already devices that claim to do more than monitor, such as headbands that alert wearers when they become distracted or wristbands that administer electric shocks to smokers who want help quitting. Electronics companies promise to transform medicine with wearables that can treat symptoms or manage care. Devices are emerging that alert people with epilepsy to incipient seizures, help prevent anxiety attacks, and enable blind people to navigate.

But the potential of wearables crucially depends on the large amounts of data they access and generate. And that leads to two problems that researchers and technology developers are struggling to solve: finding improved ways to transmit data to and from wearables, and keeping all that information safe. With everything from toasters to cars now connecting wirelessly to the Internet, demands on a finite bandwidth are rapidly straining the system. Nearly half a billion new devices started chattering over mobile broadband last year alone, pushing mobile traffic to 25 times what it was just 5 years ago. And wearables are leading to new

security concerns, from the use of highly personal data to track people's activity to maliciously attacking their online presence.

"It's a cliché that whenever there's a new technology we start talking about Huxley and *A Brave New World*, but with wearables — and what's loosely termed the Internet of Things — we truly are entering into a new era, and we have to start thinking of these issues," says Anupam Joshi, head of the Center for Cybersecurity at the University of Maryland, Baltimore County.

TRAFFIC JAM

By the end of 2014, global mobile-data traffic reached 2.5 exabytes (2.5 billion gigabytes) per month according to the networking-technology company Cisco Systems. Of that, the world's 100 million or so wearable devices were generating 15 million gigabytes of monthly traffic on what is a physically finite portion of the electromagnetic spectrum, with their number expected to increase fivefold by 2019 (see "The catch with gadgets"). On top of the surge in those devices, there will be even greater chances for gridlock, as more people start wearing headsets that deliver data-hungry virtual and augmented reality experiences, says Robert Heath, a professor in electrical engineering at the University of Texas at Austin.

All these devices clog up the airwaves, impairing performance and threatening essential internet traffic. To help ease congestion in the United States, the government pledged in 2010 to free up an extra 500 megahertz (MHz) within ten years, a doubling of the bandwidth available for mobile devices at the time. But even this is unlikely to be enough, according to a more recent report prepared for CTIA-The Wireless Association, a communications industry group based in Washington DC. It estimates that 350 MHz will need to be added from 2015 onwards to keep up with US demand by the end of 2019, 150 MHz more than the government estimate for that period. And limited bandwidth is a global problem, with each country dealing with it in its own way. In India, where users have access to just one-tenth of the bandwidth available to people in the United States, there are calls for spectrum sharing

and the freeing up of channels currently devoted to the military. In the United Kingdom, the government has approved the use of old analogue TV bandwidths; the first networks of smart devices using these frequencies could be rolled out by the end of the year.

For their part, telecom companies need to make more efficient use of the spectrum. One way is to look beyond the crowded parts of the airwaves in the radio and television bands. Data from all the wearables on one person could flow through a body-area network designed to use a completely different part of the spectrum, such as the millimetre wavelengths. Then just one device would use the more congested bands to communicate all the data to the Internet. This creates its own problems, however, because shorter wavelengths demand more power and can be blocked by people's bodies. So researchers such as Heath are trying to get around those difficulties by, for example, optimizing antennas to reduce interference and power consumption. Improvements in steerable communication beams could also lead to better ways of transmitting millimetre-wavelength signals.

Also promising is the idea of taking wireless communications into the visible-light realm using light-emitting diodes (LEDs) — which produce light and can act as photoreceptors — to communicate either between wearables or to talk directly to the Internet. Wearables that incorporate LEDs could use visible light to wrap a person in a body-area network. That would sense every movement and communicate the information to the light fittings in a room, which would be connected to the Internet through their power wiring. Although this technology relies on visible wavelengths, the signals are imperceptible. “LEDs blink so fast that the human eye cannot tell,” says Daniele Puccinelli, an electrical engineer at the University of Applied Sciences of Southern Switzerland in Manno, who studies visible-light communications.

Harald Haas, who researches mobile communications at the University of Edinburgh, UK, plans to test a visible-light system in hospitals within the next year. Patients will wear wristbands that monitor their temperature and relay the data using LEDs that communicate with the hospital's lighting.

A broader approach might have wearable devices from many people relaying information to each other rather than having each connect to the Internet. This concept underpins the multitiered networks promised by the much-vaunted fifth-generation (5G) mobile-communication systems that are predicted to be up and running in many parts of the world by 2020. In situations where crowds of people are trying to access the same content — travel information after a sports match, for instance — one device could act as a ‘seed,’ distributing the data to others in this network, which would reduce the number of times the data need to be downloaded from the Internet.

One of the most attractive approaches makes devices smarter about when and how they use communication channels. These ‘cognitive radios’ sniff out underused regions of bandwidth and opportunistically hop into those gaps, speeding up communications. To reach their optimum potential, bandwidths would need to be more open, so that devices could jump onto a licensed frequency to communicate, and then drop off the spectrum when someone with higher priority enters. Although techniques based on this principle have been used for decades, cognitive radio will take it to a new level of efficiency, with devices smart enough to negotiate with each other to divvy up the available spectrum.

Cognitive radios have great potential, but their development in the wearables realm is being held back by a lack of accepted standards and protocols for how this frequency hopping might work in practice, says Ekram Hossain, an electrical engineer at the University of Manitoba, Canada. “Until there is a standard, there won't be products,” says Hossain, who adds that the research needed to establish these standards is under way.

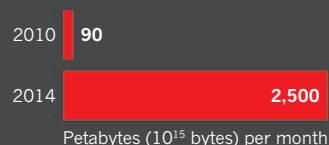
KEEPING SAFE

When 176,000 people swarmed through the Consumer Electronics Show in Las Vegas in January, some of the hottest items were the crop of new wearable devices, ranging from watches and glasses to the Pacifi-i, a smart pacifier, or baby soother, that monitors an infant's temperature

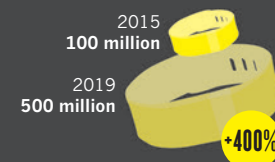
THE CATCH WITH GADGETS

Data concerns could thwart the vast expansion of wearable electronics. The number of devices is rising quickly, putting strain on the already clogged mobile network. There are also worries about security and privacy.

GLOBAL MOBILE DATA TRAFFIC

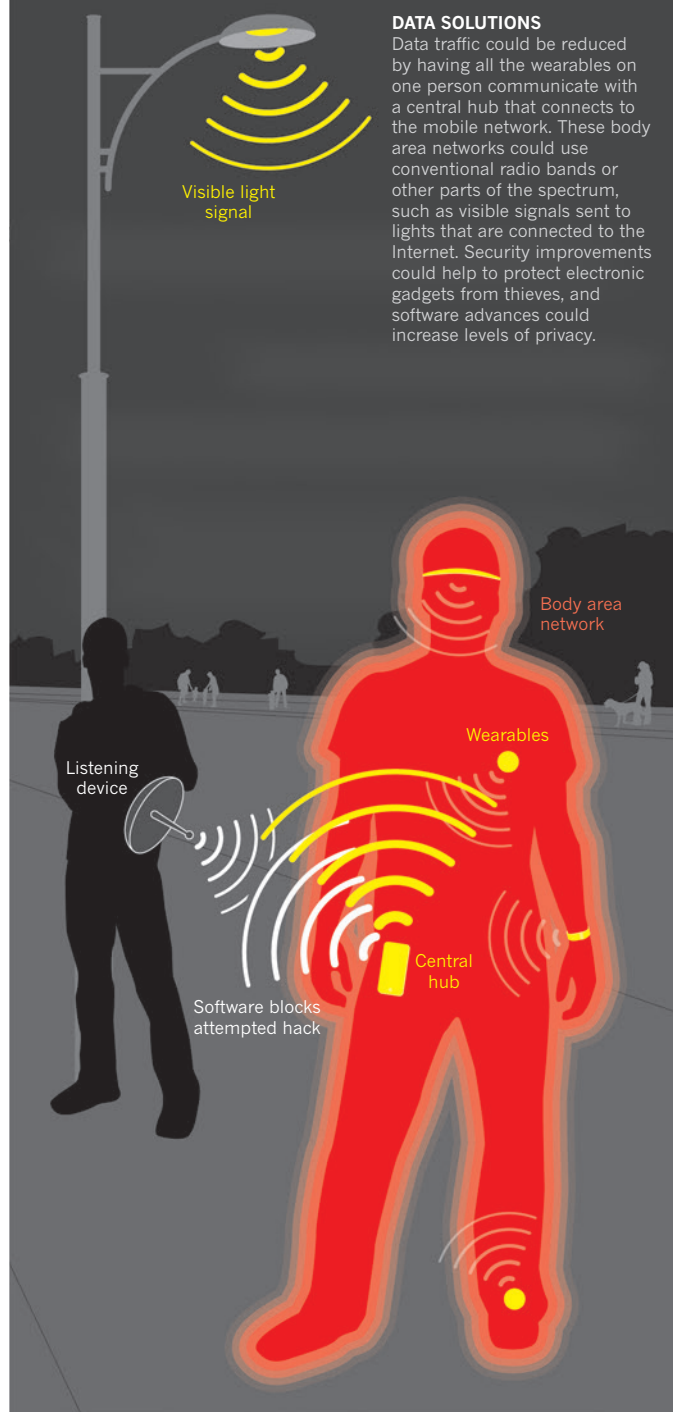


NUMBER OF WEARABLES



DATA SOLUTIONS

Data traffic could be reduced by having all the wearables on one person communicate with a central hub that connects to the mobile network. These body area networks could use conventional radio bands or other parts of the spectrum, such as visible signals sent to lights that are connected to the Internet. Security improvements could help to protect electronic gadgets from thieves, and software advances could increase levels of privacy.



and transmits the data to a parent's phone. And if those parents were stressed out, they could try the Melomind headset, which is advertised to measure the brain's electrical activity, beam it to a phone and then select the most appropriate music to help the wearer relax.

Despite all the hype about wearables, there is also considerable scepticism about the gadgets available today. "Lots of people view wearables as just toys," says Puccinelli.

But signs point to them being much more useful in the near future, particularly in the medical arena. Wearables are increasingly measuring aspects of human physiology, providing electrical stimulation to the brain and even injecting medication. These applications come with potential risks for users.

A key hurdle for the wearable revolution arises from the wealth of personal data they gather about their users. Surveys show that users worry about how these devices invade their privacy, as they upload intimate data to potentially vulnerable servers owned by companies that could change their terms of service, be bought out or go out of business.

When the Pew Research Center, an independent fact-gathering organization in Washington DC, canvassed 1,600 experts in 2014 about the future of the Internet, many expressed similar worries. "The realities of this data-drenched world raise substantial concerns about privacy and people's abilities to control their own lives," according to the report. Those concerns have been compounded by some high-profile incidents, such as when users of Fitbit activity trackers allowed their activity logs to be publicly accessible, unwittingly revealing when they had sex. When that was realized in 2011, Fitbit quickly took action to fix the problem.

In another high-profile incident, the introduction of Google Glass headsets two years ago triggered concerns that users would capture images of passers-by without their knowledge. Researchers at the Center for Cybersecurity took this opportunity to apply their work on computer codes that enforce privacy policies. They built the wryly named FaceBlock app, which blocks out the faces of people who have requested privacy from photographs taken by Google Glass. But for this to work, a Google Glass owner would have to opt in by installing the app. So the only way for such a system to reliably provide privacy would be for manufacturers to make it standard and implement it with dedicated hardware, says Joshi. "Let's say that Google was to build in a feature like this into every Google Glass so that it would automatically obey these kinds of commands — then it would work."

Security concerns go hand in hand with privacy. Although encryption is becoming more pervasive and advanced, it is sometimes not used in low-cost wearable devices. Last year, researchers at the California-based information-management company Symantec, revealed that the location of many health monitors, including some from market leaders, can be easily tracked. And some of them wirelessly communicate passwords in clear text, which makes them vulnerable to hacking. Even if a health monitor is encrypted, the smartphone or hub device that links it to the Internet could also be a weak point, either because of unnecessarily broad permissions or because of malware.

"If you're not encrypting the data you're definitely not secure," says Bogdan Carbunar, a security researcher at Florida International University in Miami. "Even if you're encrypting the data you can still not be secure." Carbunar worked with a team, including a researcher from IBM, on security holes in two popular low-cost wearable fitness devices, the Fitbit Ultra and the Garmin Forerunner. They found that

by impersonating the devices' trusted web servers, they could fool the gadgets into uploading false data — even nonsensical numbers such as millions of steps in one day (see M. Rahman *et al. IEEE Trans. Mobile Comput.* <http://doi.org/636; 2015>).

The researchers also found that they could inject data onto a tracker of their own, which would compromise data accuracy, something that could become a problem if fitness data are tied to health-insurance premiums, as they have been in some companies. Fitbit told *Nature* that it had been aware of the problem, which has been addressed in subsequent products. Garmin did not respond to requests for comment.

According to Carbunar, security adds costs for manufacturers in terms of money, development time, device size and power consumption. But researchers are pushing to minimize those costs. After

working out how to hack the devices, Carbunar and his team devised a way to keep them safe. They developed SensCrypt, an encryption protocol designed specifically for low-energy fitness trackers that reduces communications costs. It uses a procedure called symmetric key encryption to protect against remote attacks and to provide some security even if the device is stolen and tampered with. The researchers were unable to implement it on Fitbit or Garmin devices because they use closed-source code, but have tested their system on an open-source proxy.

Even with high levels of encryption, devices could still be vulnerable to attack, says Bart Preneel, a cryptographer at the KU Leuven and iMinds research centre in Belgium. Preneel specializes in understanding and preventing side-channel attacks: attempts by hackers to infiltrate mobile devices by detecting fluctuations in the power usage and using these to calculate encryption keys and other secure

information. "These attacks can be made at a distance of 10 or 20 metres," he says. This type of attack was discovered around 20 years ago in relation to bank cards, but ways of preventing it are not implemented in many wearable devices, particularly implanted medical technology.

Some companies have tried to improve security on mobile devices and wearables by equipping them with biometric devices such as fingerprint readers and iris scanners. But even these are insecure: researchers and hackers have shown how high-resolution cameras can capture someone's iris from a distance and how to steal a fingerprint using a phone's camera.

But Preneel says that biometrics are promising for encryption if designers focus on measures that are not so easy to discover. There are already wearables that authenticate users on the basis of their heartbeat pattern. In the long run, Preneel envisages using internal signals from the body, such as DNA or the internal microbial community, to pair with wearable gadgets so that the devices would unlock only when in close proximity to the owner.

With these kinds of improved security — and many upgrades in communications networks — a lost tourist in the future would stand a better chance of getting their wearables to work in a crowded plaza. Tom would easily be able to summon a map of the city on his lenses and would know his personal data were safely encrypted. Following the highlighted route, he might even make it to the station with enough time to get a coffee and charge his gadgets. It may not be the technological utopia imagined by some wearables enthusiasts, but at least he will catch his maglev. ■

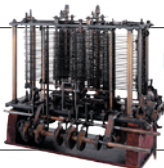
Kat Austen is a freelance writer in Berlin.

WITH EVERYTHING FROM TOASTERS TO CARS NOW CONNECTING WIRELESSLY TO THE INTERNET, DEMANDS ON A FINITE BANDWIDTH ARE RAPIDLY STRAINING THE SYSTEM.

COMMENT

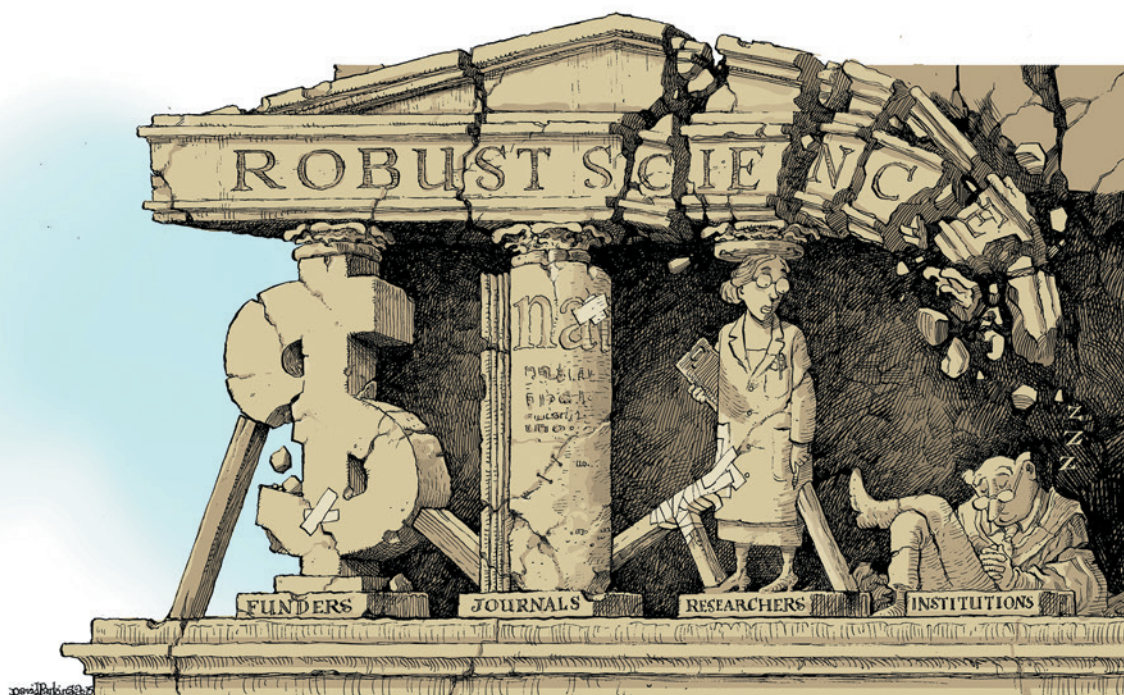
ENERGY Subsidize renewables to break fossil-fuel stronghold and spur carbon pricing **p.27**

HISTORY Richard Holmes on the scientific and cultural legacy of Ada Lovelace **p.30**



EQUALITY Memoir of a salmon run to nowhere **p.31**

CONSERVATION Offsets are often the only way to meet commitments **p.33**



Institutions must do their part for reproducibility

Tie funding to verified good institutional practice, and robust science will shoot up the agenda, say **C. Glenn Begley, Alastair M. Buchan and Ulrich Dirnagl**.

Ireproducible research poses an enormous burden: it delays treatments, wastes patients' and scientists' time, and squanders billions of research dollars. It is also widespread. An unpublished 2015 survey by the American Society for Cell Biology found that more than two-thirds of respondents had on at least one occasion been unable to reproduce published results. Biomedical researchers from drug companies have reported that one-quarter or fewer of high-profile papers are reproducible^{1,2}.

Many parties are addressing the problem. Funding bodies such as the US National Institutes of Health (NIH) have announced

training initiatives³ and explicitly instructed grant reviewers to consider whether experimental plans ensure rigour. New methods of data analysis and peer review have been proposed to deflate bias.

Several journals, including *Nature* and *Science*, have updated their guidelines and introduced checklists. These ask scientists whether they followed practices such as randomizing, blinding and calculating appropriate sample size. *Science* has also added statisticians to its panel of reviewing editors. Philanthropic and non-profit organizations have sponsored projects to improve robustness.

Funders' policies, journal guidelines and

widespread soul-searching are necessary. But they are not sufficient.

Conspicuous by their absence from these efforts are the places in which science is done: universities, hospitals, government-supported labs and independent research institutes. This has to change. Institutions must support and reward researchers who do solid — not just flashy — science and hold to account those whose methods are questionable.

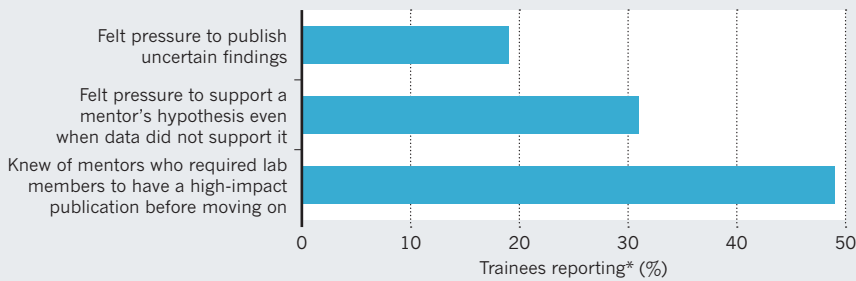
SPOT THE SHIRKERS

Although researchers want to produce work of long-term value, multiple pressures and prejudices discourage good scientific

ILLUSTRATION BY DAVID PARKINS

PRESSURED FINDINGS

A survey of US biomedical trainees suggests that the push to publish spurs unreliable results.



*Online survey of ~140 trainees at the MD Anderson Cancer Center in Houston, Texas.

► practices. In many laboratories, the incentives to be first can be stronger than the incentives to be right.

Discussions of conflicts of interest typically centre on relationships with industry, but academic scientists face more pernicious, even existential temptations. Monetary rewards are often less important than the 'currency' with which scientists advance their careers: high-level publications lead to funding opportunities, promotions, awards and other forms of recognition. These markers of scientific achievement become proxies for assessment of the work itself, and further encourage spectacular, but less than substantiated, research.

Amplifying these pressures is a human prejudice in favour of our own ideas. There is a very real temptation to ignore a result that does not conform to our preconceptions, or to recast it so that it does. Data-dredging is used to find statistically significant results that justify a publication. Sound practices such as blinding, multiple repeats, validated reagents and appropriate controls⁴ are dismissed as luxuries or nuisances.

Research institutions contribute to and benefit from these perverse incentives. They bathe in the reflected glory of their faculty; they trumpet breakthroughs published in top-tier journals, lauding achievements to the media and donors. Some even pay investigators for publications. Many require that investigators generate their salary from research grants.

An anonymous survey of around 140 trainees at the MD Anderson Cancer Center in Houston, Texas, found that nearly one-third had felt pressure to prove a mentor's hypothesis even though their experimental results did not support it, and nearly one-fifth had themselves published results they considered less than robust⁵. Nearly half knew of mentors who required lab members to publish a high-impact paper to complete training in their labs (see 'Pressured findings').

Although important, the checklists introduced by journals do nothing to shift the focus from results to the legitimacy of the process by which the results are produced. Researchers encounter these lists after they

have drawn conclusions and are ready to announce them — not when planning their research. There is no mechanism to verify that listed practices were actually employed.

The core instinct of scientists — scepticism — is punished by the current system. Institutions have a duty to reform it. They must shoulder their responsibility for training graduate students and postdoctoral fellows, for supporting the scientific behaviour of their faculty members and for the knowledge that emanates from their endeavours.

GOOD INSTITUTIONAL PRACTICE

Although there are some protections against outright fraud, few institutions have strong, transparent processes in place to discourage poor-quality science or to foster objectivity. We propose that research institutions that receive public funding should apply the same kind of oversight and support to ensure research integrity as is routinely applied for animal husbandry, biosafety and clinical work.

To conduct animal research, investigators must hold licences and undergo continuous education. Institutions appoint delegates to monitor compliance, and those delegates are held to account by regulators. Similar oversight is used for work with radioactivity and human embryonic stem cells.

These functions could be broadened to encompass established guidelines for research conduct, such as the ARRIVE (Animal Research: Reporting of In Vivo Experiments) and MIAME (Minimum Information About a Microarray Experiment) guidelines, and data sharing as required by the NIH and the National Science Foundation.

Standards already exist that define good laboratory practice to test chemicals for toxicity, good manufacturing practice and good clinical practice. These systems were introduced to ensure a degree of consistency, quality and integrity. Procedures

"Most institutions will not make the necessary move unless forced."

are in place to ensure compliance.

The scientific community should come up with a similar system for research, which we term good institutional practice (GIP). If funding depended on a certified record of compliance with GIP, robust research would get due recognition.

At a minimum, GIP should consist of the following tenets.

Routine discussion of research methods.

Many labs already comb through data and methods as a group before submitting a paper. Such discussions should be broadened and formalized across an institution. Regular department and cross-department meetings should be established to dissect manuscripts in preparation. Methods and processes (rather than conclusions) would be debated just as a competitor's paper might be critiqued in a journal club. Primary research material would be available. This practice is roughly analogous to the 'Morbidity and Mortality' conferences routine in hospitals, in which working hours are also intense.

Regular critique sessions help scientists to learn to defend their science without feeling defensive. Investigators publicly hold each other to account, and trainees learn what to demand of their own research. Anxieties can be raised informally, highlighting institutional weaknesses and systematic errors. The practice also puts a short-term focus on what has traditionally been a long-term reward: a reputation for careful science.

Reporting systems. Also well-established in clinical medicine is a system to anonymously flag occurrences that did or could have jeopardized a patient's care. Such systems are often the only way workers dare to raise concerns and admit mistakes. Similarly colleagues, graduate students and postdocs should be able to discuss concerns about sloppy science without jeopardizing their careers. Designated co-mentors, a departmental ombudsman or existing university offices of research integrity could be charged with providing a forum for informal, confidential discussions. Any formal reports should be investigated in a balanced and impartial way.

Training and standards. Some sloppiness stems from ignorance. Many investigators determine whether trainees are ready to move on by gauging the number and impact factors of their publications; instead, supervisors should base such decisions on whether their lab members understand research methods and process. Compulsory institutional training should ensure a common understanding of rigorous experimental design, research standards and objective evaluation of data. Faculty members and trainees should demonstrate their ability to spot problems such as 'cherry picking' data to make the best

story. Compliance with research standards, including data-sharing, should be supported, audited and acknowledged.

Records and quality management. Laboratory notebooks and records must be available for independent review. Electronic laboratory notebooks facilitate collaboration, supervision and record keeping, and can link records to the original data. One of our institutions (U.D.'s) is now adopting these system-wide. Random audits should be conducted to guarantee that experimental data are duly recorded and that elements of good research practice are routine. Such spot-checks are commonplace in industry.

Appropriate incentive and evaluation systems. Institutions should find ways to deter non-compliance with guidelines, poor mentoring and scientific sloppiness. Faculty members with poor records should face loss of laboratory space and trainees, decreased funding and potential demotion. Conversely, faculty members who excel as mentors and careful experimentalists should be rewarded. Appropriate metrics should be developed so that promotions are based on robustness and high-quality mentoring, rather than simply on high-profile publications⁶. Surveys such as that conducted at MD Anderson exemplify one way in which administrators can gain the insight necessary to improve the research environment. Institution-level metrics could help to monitor overall performance and remind all researchers and administrators of their responsibility to the scientific community.

Enforcement. Institutions should investigate egregious lapses and record them in a routine, transparent way. Departments of research integrity or other centres of excellence should be funded, staffed and given enough authority to prevent, detect, investigate and penalize poor-quality research. They should also be charged with promoting an institutional culture that nurtures robustness.

GETTING TO GIP

The systems needed to promote reproducible research must come from institutions — scientists, funders and journals cannot build them on their own. These kinds of changes will require additional money, infrastructure, personnel and paperwork. The load on institutions and investigators will be real, but so is the burden of irreproducible research. Even if it is accompanied by an apparent decrease in productivity, the resulting increase in research quality will be well worth the costs.

Still, most institutions will not make the necessary moves unless forced. Funding bodies should make GIP a prerequisite for receiving a grant. The concept has gained some traction: last year, Science Foundation Ireland announced plans to conduct external audits on some of the labs that it supports.

There will not be one ideal solution. Faculty members, trainees and administrators will need to come together for honest, difficult discussions to restructure institutions. Neither scientists nor institutions should engage in mere box checking; new practices must restrain sloppiness while interfering only minimally with the many scientists who are behaving well.

Large-scale change is possible. In the 1970s, clinical research had little rigour or oversight. Now clinical trials routinely include concurrent control groups, double-blinding, pre-specified endpoints, power calculations to determine patient numbers and analysis plans that thwart bias. In addition, primary data are available for independent statistical analysis by regulatory authorities. At the time, these changes were controversial; many physicians believed them to be unnecessary and regressive.

Nothing an institution can do will prevent misconduct altogether. This is not the goal. Rather, it is to support the work of well-meaning scientists, to reduce the waste from biased results, and to relieve some of the pressures that encourage sloppy science. ■

C. Glenn Begley is chief scientific officer at TetraLogic Pharmaceuticals, Malvern, Pennsylvania, USA. **Alastair M. Buchan** is dean of medicine and head of the Medical Science Division, Oxford NIHR Biomedical Research Centre, University of Oxford, UK. **Ulrich Dirnagl** is director of the Center for Stroke Research at the Universitätsmedizin Charité Berlin, Germany.
e-mail: cgbegley@tlog.com

1. Begley, C. G. & Ellis, L. M. *Nature* **483**, 531–533 (2012).
2. Prinz, F., Schlange, T. & Asadullah, K. *Nature Rev. Drug Discov.* **10**, 712–713 (2011).
3. Collins, F. S. & Tabak, L. A. *Nature* **505**, 612–613 (2014).
4. Begley, C. G. *Nature* **497**, 433–434 (2013).
5. Mobley, A., Linder, S. K., Braeuer, R., Ellis, L. M. & Zwilling, L. *PLoS ONE* **8**, e63221 (2013).
6. Ioannidis, J. P. & Khoury, M. J. *J. Am. Med. Assoc.* **312**, 483–484 (2014).

Push renewables to spur carbon pricing

Make wind and solar power even cheaper by opening up access to the electricity grid and ending fossil-fuel subsidies, urge **Gernot Wagner** and colleagues.

Putting a price on carbon dioxide and other greenhouse gases to curb emissions must be the centrepiece of any comprehensive climate-change policy. We know it works: pricing carbon creates broad incentives to cut emissions. Yet the current price of carbon remains much too low relative to the hidden environmental, health and societal costs of burning a tonne of coal or a barrel of oil¹. The global average price is below zero, once half a trillion dollars of fossil-fuel subsidies are factored in.

Momentum towards effective carbon pricing is building. California, joined by the Canadian province of Quebec, leads by pricing 85% of such emissions at around US\$12 per tonne. Sweden applies the highest value globally on half of its carbon dioxide emissions, at up to \$125 per tonne. The European Union has the largest system in terms of tonnes covered, pricing 45% of its greenhouse-gas emissions at about \$8 per tonne. China is experimenting with regional cap-and-trade systems. And the US Clean Power Plan encourages states

to meet emissions-reduction targets through market-based mechanisms. Yet global emissions continue to climb.

The current inadequacy of carbon pricing stems from a catch-22. Policymakers are more likely to price carbon appropriately if it is cheaper to move onto a low-carbon path. But reducing the cost of renewable energies requires investment, and thus a carbon price.

In our view, the best hope of ending this logjam rests with tuning policies to drive down the cost of renewable power ►



China (left) has invested in renewables supply, whereas Germany (right) has subsidized demand.



► sources even further and faster than in the past five years. The cost of crystalline silicon photovoltaic (PV) modules has fallen by 99% since 1978 and by 80% since 2008; installation costs for wind power have also dropped, and solar and wind capacity has grown² (see “The rise of renewables”). Prices will continue to fall, but — without more help — the decrease will not be fast enough to make a dent in the climate problem.

Some obstacles are technological; many are policy-driven^{3,4}. Most energy regulations were set with the fossil-fuel industry in mind, and energy providers fight to preserve their existing assets rather than adapt to new conditions. More strategic coordination of energy resources, grid operation and climate policy is needed, keeping in mind trade-offs.

GRID OF POWER

We call for policymakers to modernize and open up access to power grids, and to subsidize key technologies — particularly for storage. Renewables must have the same access to the grid as fossil sources; and grids must accept and manage distributed generation and intermittent flows. Other investments must include support for research, development and demonstration of energy storage and new low-carbon energy technologies. Barriers to international trade in renewable technologies and services must be lowered.

We are in the middle of a low-carbon-

energy revolution. Germany has proved an early driving force on the demand side and China has been strong on the supply side. Germany’s Renewable Energy Sources Act, adopted in 2000, guaranteed 20 years of grid access and fixed prices for its solar- and wind-power producers. German electricity consumers are subsidizing the expensive early stages of the development, deployment and integration of renewables to the tune of more than \$20 billion a year. In 2014, despite the country exporting more electricity than ever to its neighbours and phasing out nuclear power, carbon emissions from the German power sector were the second lowest since 1990.

Meanwhile, China’s climate, energy and industrial policies have boosted the manufacturing scale of renewable technologies, expanding solar PV production more than 100-fold since 2005 (ref. 5). As a result, PV-module prices have come down faster than anticipated. Other countries are taking note. More than half of US states are mandating an increase in the proportion of renewable power and have an incentive to expand such programmes under the Clean Power Plan.

“With no single ‘best’ solution available, controlled policy experiments are needed.”

But ‘business as usual’ is not enough. Even Germany — where solar energy meets more than 50% of national electricity demand on a sunny Sunday afternoon (when the sun is out and demand is low) — gets more than half of its annual electricity from coal and natural gas. Further reduction of fossil fuels relative to renewables is not assured. Fossil-fuel prices are volatile, and demand for renewables stalls when coal and natural gas are cheap.

Poorly designed subsidies can be counter-productive because some low-carbon technologies perform better than others. Some forms of bioenergy may increase rather than reduce net emissions, owing to energy-intensive, fossil-fuel-based production processes and land-use changes, such as deforestation⁶. The reservoirs of hydropower dams may leak methane, and nuclear plants are expensive and carry large potential environmental risks. Still, the worst offenders are subsidies for conventional fossil fuels.

OPEN EXPERIMENT

The ideal solution is to vary the price of electricity by time and location, reflecting the full costs of generation and distribution — including environmental costs. But that leads to another dilemma: proper pricing at all levels is politically and analytically difficult. Compromises and alternative instruments are needed. For example,

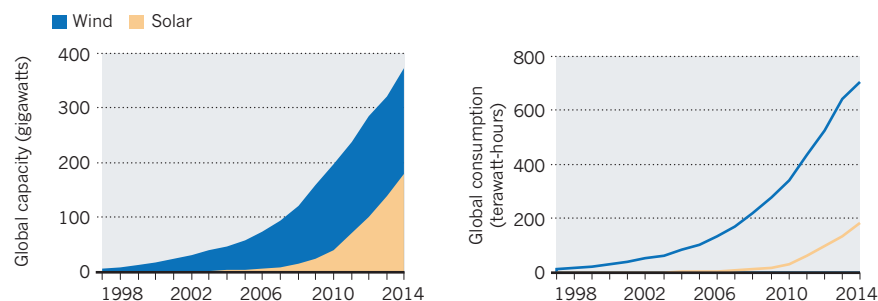
CHINA: XINHUA NEWS AGENCY/EVEVINE; GERMANY: WERNER DIETERICH/WESTEND61/CORBIS



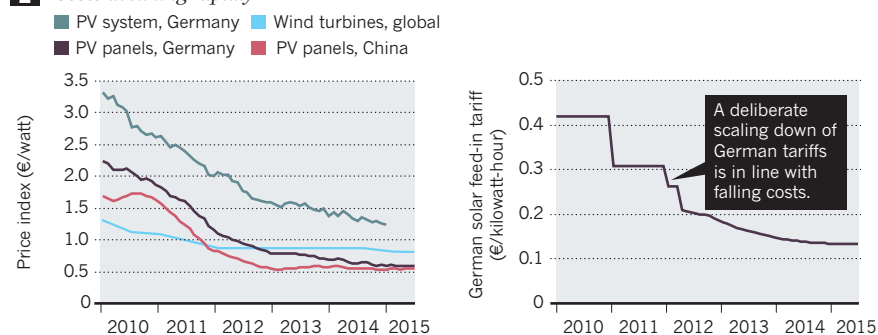
THE RISE OF RENEWABLES

Global wind- and solar-power capacities have grown by 40–50 gigawatts each year since 2008, with consumption also rising (1). Meanwhile, prices of photovoltaic (PV) panels and solar energy have fallen steeply since 2010 (2), in part driven by climate and energy policies and more-efficient manufacturing.

1 Consumption and capacity increasing



2 Costs declining rapidly



German feed-in tariffs that guarantee fixed prices for renewable-energy generation led to large increases in solar and wind installations. However, as the prevalence of renewable energies increases, the system needs — and is undergoing — reform. With no single ‘best’ solution available, controlled policy experiments are needed.

First, policymakers must check that interventions pass the benefit–cost test. Given how far the world remains from a sensible global climate policy, this is often a low threshold. Many direct subsidies that support renewables — especially solar energy — are beneficial, not least because they spur learning-by-doing⁷.

Second, any renewables policy should make a national — and eventually global — carbon cap or tax more likely. If an intervention might derail such efforts, then stop. If it paves the way for stronger climate policy, try it. The Clean Power Plan, for example, encourages flexible, market-based ways of achieving emissions-reduction goals and creates a framework for trading between states — a clear boon to sensible carbon pricing.

Third, governments should break up non-competitive arrangements around grid access. Funding and regulation should support the modernization of power grids to allow new renewable energy sources to be integrated. So that everyone pays their fair share towards the upkeep of the infrastructure, grid users should be charged — but

with caution. High rates might provoke some consumers to disconnect, increasing costs for the rest. Boosting supply of renewables during peak times might in turn lower peak pricing and thus (perversely) decrease overall incentives for renewables adoption⁸.

Fourth, the energy sector should be viewed in its entirety. For instance, increased electric-vehicle use could spread electricity demand more evenly throughout the day, flattening traditional peaks. It would also help to lower the prices of battery technologies, hastening systemic change in the transport and electricity sectors.

Ambitious renewables policies should be followed by strengthened climate policies. For example, rapid renewables deployment has reduced Germany’s carbon emissions but has not brought down the EU total, because German emissions are capped under the EU’s Emissions Trading System. The decrease in Germany, all else being equal, is compensated by emissions increases elsewhere under the cap. All else must not be equal. The cap ought to be tightened.

These are the sorts of pieces that need to come together to deepen solar and wind penetration levels and achieve the ‘holy grail’ of climate policy: an effective carbon price. ■

Gernot Wagner is lead senior economist at the Environmental Defense Fund in Boston, Massachusetts, USA, and adjunct associate professor of international and public affairs

at Columbia University, New York, USA.

Tomas Käberger is professor of industrial energy policy at Chalmers University of Technology, Gothenburg, Sweden.

Susanna Olai is an environmental economist at the University of Gothenburg, Sweden.

Michael Oppenheimer is professor of geosciences and international affairs in the Woodrow Wilson School and the Department of Geosciences, Princeton University, New Jersey, USA.

Katherine Rittenhouse is an economic analyst at the Environmental Defense Fund in Boston, Massachusetts, USA.

Thomas Sterner is professor of environmental economics at the University of Gothenburg, Sweden, and senior contributing economist at the Environmental Defense Fund in New York, USA.

e-mail: gernot@gwagner.com

1. Wagner, G. & Weitzman, M. L. *Climate Shock: The Economic Consequences of a Hotter Planet* (Princeton Univ. Press, 2015).
2. Trancik, J. E. *Nature* **507**, 300–302 (2014).
3. Prentiss, M. *Energy Revolution: The Physics and the Promise of Efficient Technology* (Harvard Univ. Press, 2015).
4. Energy Initiative, Massachusetts Institute of Technology. *The Future of Solar Energy* (MIT, 2015).
5. Mathews, J. A. & Tan, H. *Nature* **513**, 166–168 (2014).
6. Searchinger, T. D. *et al. Science* **326**, 527–528 (2009).
7. van Benthem, A., Gillingham, K. & Sweeney, J. *Energy J.* **29**, 131–151 (2008).
8. Darghouth, N. R., Wiser, R., Barbose, G. & Mills, A. *Net Metering and Market Feedback Loops* (Lawrence Berkeley National Laboratory, 2015).

SOURCES: 1. BP STATISTICAL REVIEW OF WORLD ENERGY; 2. PVCHANGING (GERMAN, CHINESE PV PANELS); PHOTOVOLTAIK-GUIDE.DE (GERMAN PV SYSTEM); BLOOMBERG NEW ENERGY FINANCE (GLOBAL WIND TURBINE); BUNDESNETZAGENTUR (GERMAN SOLAR FEED-IN TARIFF)

COMPUTER SCIENCE

Enchantress of abstraction

Richard Holmes re-examines the legacy of Ada Lovelace, mathematician and computer pioneer.



Ada Lovelace, painted in 1835 by Margaret Sarah Carpenter.

The bicentenary of Augusta Ada King, Countess of Lovelace, heralds the critical reassessment of a remarkable figure in the history of Victorian science. Ada Lovelace (as she is now known) was 27 years old and married with 3 children when she published the first account of a prototype computer and its possible applications in 1843. Her 20,000-word paper was appended as seven Notes to a translation of a descriptive article, *Sketch of the Analytical Engine Invented by Charles Babbage, Esq.*

Lovelace's account was the fruit of one of the most intriguing collaborations in the annals of science: her friendship with Charles Babbage, Lucasian Professor of Mathematics at the University of Cambridge, UK, and inventor of the landmark analytical engine. The Notes eventually brought Lovelace both acclaim and notoriety. Babbage himself described her unforgettably to the physicist Michael Faraday as "that Enchantress who has thrown her magical spell around the most abstract of Sciences and has grasped it with a force that few masculine intellects (in our own country at least) could have exerted over it".

The exact nature of that force and enchantment continues to puzzle historians of science, not least because Lovelace's correspondence, largely archived at the Bodleian Library in Oxford, has not been fully published (see selections by Dorothy Stein in *Ada* (MIT Press, 1985) and Betty A. Toole in *Ada, Enchantress of Numbers*; Strawberry, 1992). What has emerged is the hitherto unsuspected range of Lovelace's interests and contacts, which linked the worlds of Victorian science and literature.

Lovelace was the only legitimate child of the poet Lord Byron. She never met her father, self-exiled in Italy and Greece, but inherited much of his rebellious spirit and something of his unstable genius. She directed it towards science, declaring: "I do not believe that my father was (or ever could have been) such a Poet as I shall be an Analyst (& Metaphysician); for with me the two go together indissolubly".

She was brought up with pathological severity by her mother, the brilliant Lady Annabella Byron — dubbed "the Princess of Parallelograms" for her own fascination with mathematics — and a squadron of female advisers whom Lovelace christened the Furies. Forbidden to read her father's poetry, young Ada was encouraged to study mathematics, astronomy and music, and allowed to design flying machines, play the harp and commune with her cat, Puff. In her early twenties she began to study the new calculus under Augustus De Morgan, a proponent of Boolean algebra, who described her as potentially more promising than any 'senior wrangler', or first-class Cambridge maths student.

In spring 1834, Lovelace met her first great

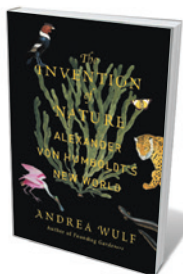
mentor: Mary Somerville, the translator of French astronomer Pierre-Simon Laplace and author of the groundbreaking popular study *On The Connexion of the Physical Sciences* (1834). Somerville demonstrated that women could make their mark in science (R. Holmes *Nature* **514**, 432–433; 2014). It was she who introduced Lovelace to Babbage at one of his champagne-and-science receptions in Marylebone, London, where Charles Darwin, astronomer John Herschel and geologist Charles Lyell were frequent guests. At these soirées, Babbage displayed a model of his early difference engine — a brass calculating machine capable of tabulating higher-order polynomial functions — alongside a silver automaton in the form of a dancing ballerina. Most guests were drawn to the ballerina; Lovelace, Babbage noticed, was entranced by the gleaming cogs of the calculating machine. Thus the unlikely friendship began.

When Ada married William King, later Earl of Lovelace, in 1835, her London town house brought her even closer to Babbage. Their mathematical correspondence, both serious and teasing, focused on the analytical engine and the possibilities of mathematical and symbolic calculation. Thus in 1840 Lovelace was discussing the elimination game solitaire, in which 26 marbles must ‘jump’ each other, in an apparently unpredictable sequence, until only one remains. She challenged Babbage to consider whether there could be “a mathematical formula ... on which the solution depends, and which can be put into symbolical language”. She added, “Am I too imaginative for you? I think not.”

By 1841 Lovelace was developing a concept of “Poetical Science”, in which scientific logic would be driven by imagination, “the Discovering faculty, pre-eminently. It is that which penetrates into the unseen worlds around us, the worlds of Science.” She saw mathematics metaphysically, as “the language of the unseen relations between things”; but added that to apply it, “we must be able to fully appreciate, to feel, to seize, the unseen, the unconscious”. She also saw that Babbage’s mathematics needed more imaginative presentation. So when a scientific paper on the analytical engine was published by Italian engineer Luigi Menabrea, Lovelace (perhaps inspired by Somerville’s translation of Laplace) translated it from the original French. A delighted Babbage encouraged her to add a commentary. When published in the British journal *Scientific Memoirs* (volume 3, October 1843), Lovelace’s ‘translator’s Notes’ had expanded to twice the length of Menabrea’s paper, and were much more far-reaching. This is the work that eventually made both the engine and Lovelace famous. ▶

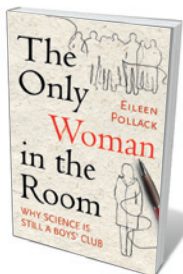
➔ **NATURE.COM**
For more on science
in culture see:
[nature.com/
booksandarts](http://nature.com/booksandarts)

Books in brief



The Invention of Nature: Alexander von Humboldt's New World *Andrea Wulf* KNOPF (2015)

Alexander von Humboldt (1769–1859) electrified fellow polymaths such as Johann Wolfgang von Goethe, discovered climate zones and grasped the impact of industrialization on nature. In her coruscating account, historian Andrea Wulf reveals an indefatigable adept of close observation with a gift for the long view, as happy running a series of 4,000 experiments on the galvanic response as he was exploring brutal terrain in Latin America. Most presciently, and at a time of fragmenting disciplines, he saw life as a “net-like intricate fabric” and brilliantly synthesized the sciences in his grand treatise *Cosmos*.



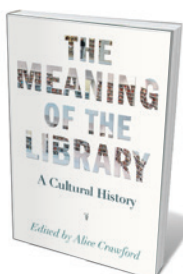
The Only Woman in the Room: Why Science Is Still a Boys' Club *Eileen Pollack* BEACON (2015)

In the 1970s, Eileen Pollack was one of the first women to earn a bachelor's degree in physics at Yale University in New Haven, Connecticut. Isolated and unencouraged, she abandoned dreams of a life in cosmology and turned to writing. In this investigative memoir, Pollack uses her own experience and interviews with students and academics as a lens on gender and science. Many will wince in sympathy over the biases and sexism that made Pollack's academic career a salmon run to nowhere, yet despite ongoing inequalities in physics, she senses hopeful shifts in awareness.



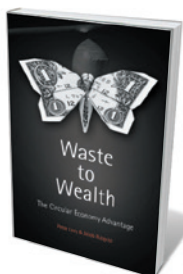
Weatherland: Writers and Artists Under English Skies *Alexandra Harris* THAMES & HUDSON (2015)

This is a gorgeous piece of writing, sure to grip bibliophiles and the meteorologically inclined alike. Scouring English art and literature for references to weatherscapes, Alexandra Harris has magicked them into a subtle meditation on the nation's changeable culture. Snippets of science intersperse discussions of Shakespeare's tempestuous dramas, the “gothic fogs” of Charles Dickens's 1853 *Bleak House*, the rain-soaked revelations of poet Ted Hughes and more. Harris captures the evanescent interplay of mind and sky, just as climate change could muddy that relationship out of all recognition.



The Meaning of the Library: A Cultural History *Editor Alice Crawford* PRINCETON UNIVERSITY PRESS (2015)

The current pressures on libraries give a poignant edge to this chronicle, edited by research librarian Alice Crawford, which offers rarefied glimpses of the institution through time. Historian Andrew Pettegree reveals that printing contributed to the Renaissance library's decline; academic librarian Robert Darnton relates how eighteenth-century booksellers went through hell and high mountain passes to transport their wares; and English-literature professor Laura Marcus surveys libraries in films such as Alain Resnais's 1956 *All the Memories of the World* and Orson Welles' 1941 *Citizen Kane*.



Waste to Wealth: The Circular Economy Advantage *Peter Lacy and Jakob Rutqvist* PALGRAVE MACMILLAN (2015)

In this crisply lucid primer on the innovative sustainable-business model called the circular economy, Peter Lacy and Jakob Rutqvist make a business case for repurposing wasted resources, life cycles and embedded values such as unrecovered energy. They sketch in the historical background; discuss worked examples of business models such as the circular supply chain; describe the creation of “circular advantage”; and map out strategies for making the shift to full sustainability. **Barbara Kiser**

► Lovelace is sometimes loosely described as the first computer programmer. She did produce an elegant set of tables showing how the engine could calculate Bernoulli numbers, but based on equations supplied by Babbage. Lovelace's originality lay in her conceptual definitions of the engine's mathematical functions, and her brilliant speculations on its design possibilities, going far beyond anything Babbage himself articulated. She wrote: "We may say most aptly that the Analytical Engine weaves algebraic patterns just as the Jacquard-loom weaves flowers and leaves."

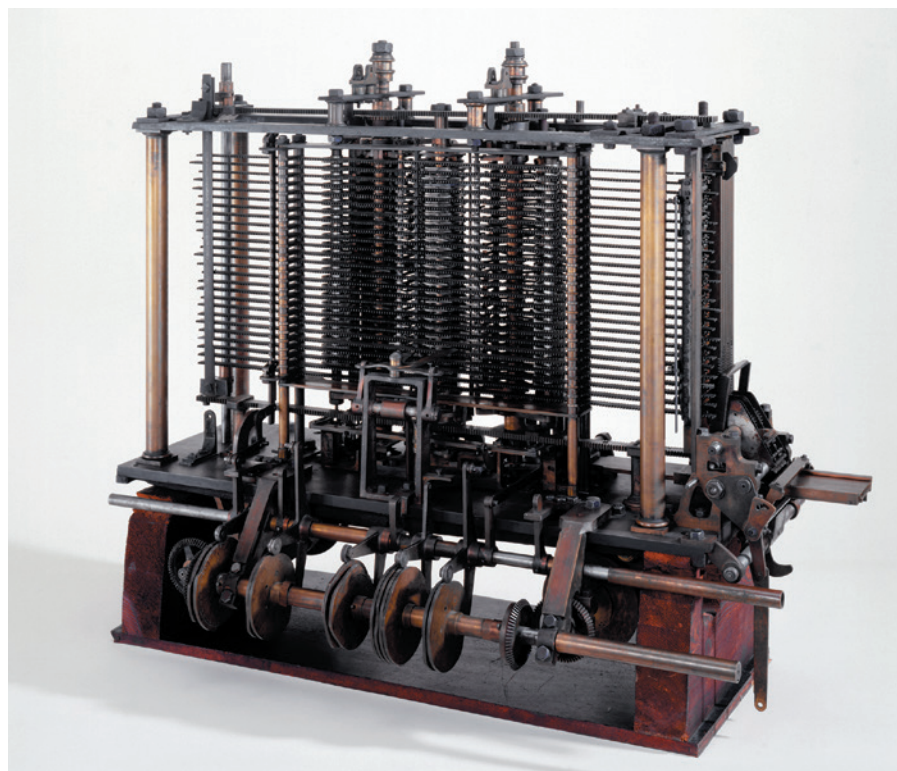
She distinguished it sharply from the difference engine and other "mere 'calculating machines'", writing prophetically that it "holds a position wholly its own... A new, a vast, and a powerful language is developed for the future". Like the Jacquard loom, it used paper punchcards that could program variable settings into a mechanical processor to be constructed from thousands of brass numerical cogs, vertically mounted in a system of calculating 'barrels', with 'loops and conditional branchings' built in. Despite having no defined power source, it was essentially the first genuine design for a working computer.

Next, Lovelace pointed out its revolutionary potential to handle purely symbolic notations, which gave it the potential to win games or compose music: "Supposing, for instance, that the fundamental relations of pitched sounds in the science of harmony and of musical composition were susceptible of such expression and adaptations, the Engine might compose elaborate and scientific pieces of music of any degree of complexity or extent."

Finally, she raised the question of whether the engine could think, but concluded that it "has no pretensions whatever to originate any thing". This was to have huge resonance. In his 1950 paper 'Computing machinery and intelligence', mathematician Alan Turing listed nine potential objections to the possibility of artificial intelligence (A. Turing *Mind* 59, 433–460; 1950). The sixth was "Lady Lovelace's Objection" that a machine cannot do anything new; he initially agreed, then wondered about it. "A better variant of the objection says that a machine can never 'take us by surprise'... Machines take me by surprise with great frequency."

So does Lovelace. As her correspondence is gradually published, the extent of her scientific interests is emerging; they included railways, experimental telegraphy,

"Lovelace's originality lay in her definitions of the engine's functions, and her brilliant speculations on its design possibilities."



A portion of Charles Babbage's never-completed analytical engine, with printing mechanism.

magnetism, animal intelligence, probability theory and photography. Along with this is the multitude of luminaries whose work she knew, from Faraday to inventor Charles Wheatstone, engineer Isambard Kingdom Brunel, social theorist Harriet Martineau and novelist Charles Dickens.

Lovelace planned to draft many other scientific papers, for example on consciousness. These ambitions, latterly fuelled by opium, became increasingly visionary: "My own great scientific object... is the study of the Nervous System, and its relations with the more occult influences of nature." She wished to become "a Newton for the Molecular Universe" of the mind. Another paper was to be on the revolutionary field theories of Faraday, to whom she wrote boldly: "I mean (unless you discourage me) to undertake your Researches for review, or at any rate as my hinge and centre for an Electrical Article." By her thirties, aware of her own celebrity, Lovelace became increasingly provocative, speaking out on science, sex and life after death — "subjects few men and no women venture to touch upon", as her father's old friend, politician John Hobhouse, observed.

Lovelace died in great pain at 36, from uterine cancer. Almost her last independent act was visiting the Great Exhibition of 1851 with Babbage, to revel in scientific and technological advances — although both knew that the analytical engine would never be built in their own lifetimes. Both poetry and science attended her deathbed: Dickens came to read from his 1848 *Dombey and*

Son; Lovelace wrote a sonnet on Newton's rainbow for her own tomb.

Lovelace's long-term influence has been as much cultural as scientific. She may have partly inspired Princess Ida in Alfred Tennyson's poem about a university of women, *The Princess* (1847). The precocious nineteenth-century teenager Thomasina Coverly in Tom Stoppard's 1993 play *Arcadia*, who understands chaos theory before it is established, was based on her. She is central to at least two science histories — James Gleick's *The Information* (Pantheon, 2011) and Walter Isaacson's *The Innovators* (Simon & Schuster, 2014) — as well as to Sydney Padua's hilarious (but scholarly) graphic novel, *The Thrilling Adventures of Lovelace and Babbage* (Particular, 2015).

Since 2009, an international Ada Lovelace Day has been celebrated every October, to promote women in science. The annual Lovelace Medal is awarded by the academy of the British Computer Society. A major academic conference will be dedicated to her work at the Mathematics Institute in Oxford this December (see go.nature.com/sbcojl). As Lovelace once wrote to a startled Faraday: "I would not miss a possible opportunity of being... useful to Science (Science whose Bride I am)!" ■

Richard Holmes is the author of *The Age of Wonder*, which won the 2009 Royal Society Prize for Science Books, and *Falling Upwards*.

e-mail: richard.holmes@osb.me.uk

Correspondence

Monitor ecosystem services from space

We suggest that Earth observation should be used to monitor ecosystem services in the run-up to implementation of the United Nations Sustainable Development Goals (SDGs; see also A. K. Skidmore *et al. Nature* **523**, 403–405; 2015).

A reliable Earth-observation framework would provide long-term spatial indicators of ecosystem services. It could capture changes in environmental and socio-economic features, for example by comparison with more than 40 years of Landsat satellite data and with information from new sensors such as the Sentinel fleet.

Earth observation could inform management decisions about how to resolve conflicting objectives that arise from the SDGs (see, for example, R. Bosch *et al. Nature* **523**, 526–527; 2015). It would help in evaluating remote effects (teleconnections) for ecosystem-service provision and usage, such as whether biofuel production in one place creates biodiversity loss, pollution or deforestation elsewhere (see J. Liu *et al. Science* <http://doi.org/627>; 2015).

We need new forms of data integration and case-study synthesis. Earth-monitoring systems must be developed with input from environmental and social scientists to link up with existing knowledge, for example by relating ecosystem services to biodiversity.

Anna F. Cord, Ralf Seppelt
Helmholtz Centre for Environmental Research (UFZ), Leipzig, Germany.
Woody Turner NASA,
Washington DC, USA.
ralf.seppelt@ufz.de

Biodiversity on canal route already at risk

The Nicaraguan government is reviewing an environmental and social impact study for a proposed 300-kilometre canal to connect the Pacific and Atlantic

oceans. As members of the specialist team who contributed to the 'baseline' biodiversity assessment for the study, we are in a position to respond to critics of the proposal (see, for example, A. Meyer and J. A. Huete-Pérez *Nature* **506**, 287–289; 2014).

The internationally recognized environmental consulting firm ERM was commissioned to produce the study. Our impression is that ERM's dealings with its local counterparts, the Nicaraguan government and the company that owns the canal concession have been mutually transparent and professional.

Contrary to the depiction of the proposed canal route by Meyer and Huete-Pérez as a pristine wilderness, human impacts are strongly evident over its entire length, particularly from agriculture. This includes nationally and internationally protected areas and Lake Nicaragua, where several fish species are already in decline (T. B. Thorson *Fisheries* **7**, 2–10; 1982; and M. T. McDavitt *Shark News* **14**, 5; 2002).

We share many of the authors' concerns for environmental integrity and biodiversity along the proposed canal route. However, there were huge losses to these even before the canal project began, and this needs to be factored into the discussion.

Jeffrey K. McCrary* *Nicaraguan Foundation for Integral Community Development, Managua, Nicaragua.*
jmccrary2@yahoo.com

**On behalf of 4 correspondents (see go.nature.com/nu2cjj for full list).*

Offsets: factor failure into protected areas

Martine Maron and colleagues assume that a nation's commitment to establishing protected areas of biodiversity provides a suitable baseline for determining the "additional" of any offset initiative based on habitat protection (*Nature* **523**, 401–403; 2015). The evidence

indicates otherwise.

A more realistic baseline would factor in the high probability that national biodiversity commitments will not be fulfilled (see M. Walpole *et al. Science* **325**, 1503–1504; 2009). For example, national conservation commitments can be overridden by development commitments.

Documented trends and local conditions should be used to establish a baseline. Carbon offsets, for example, commonly derive baselines from historical average deforestation (see go.nature.com/rvdx3x). These baselines are typically revised every ten years.

We also disagree that developing countries should withdraw from the Convention on Biological Diversity (CBD) if they are unable to fund protected areas, because that would stop them engaging with other CBD targets. Moreover, honest accounting of offset benefits must occur at the local, regional and landscape levels where conservation is accomplished.

What is most needed in offset programmes is better enforcement, so that they do not become a 'licence to trash' (see A. Villaroya *et al. PLoS ONE* **9**, e107144; 2014).

Joseph M. Kiesecker *The Nature Conservancy, Fort Collins, Colorado, USA.*

Bruce McKenney *The Nature Conservancy, Charlottesville, Virginia, USA.*

Peter Kareiva *University of California, Los Angeles, USA.*
jkiesecker@tnc.org

Galaxy γ -ray signal was not oversold

We argue that Jan Conrad's depiction of our preprint (<http://arxiv.org/abs/1503.02320>; 2015) as a case study in 'crying wolf' lacks accuracy and credibility (*Nature* **523**, 27–28; 2015).

Based on public data from NASA's Fermi Large Area

Telescope (LAT), we reported a γ -ray signal from the dwarf galaxy Reticulum II. Conrad characterizes our work as "the latest dark-matter discovery claim" and criticizes the "misuse" of public data at a time when an update from the Fermi collaboration "was imminent".

Nowhere do we claim to have discovered dark matter. Rather, our paper is devoted to quantifying the probability that the observed signal is due to random fluctuation. Our closing paragraph says "it would be premature to conclude [the signal] has a dark matter origin", then identifies future work necessary to establish such a discovery.

Our use of public data is concordant with the principles of 'reproducibility' Conrad invokes. Nevertheless, he compares our work unfavourably to a paper by the Fermi-LAT and Dark Energy Survey (DES) collaborations, who calculate a larger probability of background fluctuation (see preprint at <http://arxiv.org/abs/1503.02632>; 2015 and A. Drlica-Wagner *et al. Astrophys. J.* **809**, L4; 2015). Conrad did not disclose that he was initially an author on their submitted paper. He states that the Fermi-LAT/DES result is based on "more comprehensive re-analysis of the same data"; however, theirs is a separate analysis of different data that were released 15 weeks after both papers appeared, preventing confirmation of their results in the interim. Meanwhile, only our result was reproducible (see, for example, D. Hooper and T. Linden, preprint at <http://arxiv.org/abs/1503.06209>; 2015).

Moreover, our findings are now published in the peer-reviewed literature (*Phys. Rev. Lett.* **115**, 081101; 2015).

Alex Geringer-Sameth*
Carnegie Mellon University, Pittsburgh, Pennsylvania, USA.
alexgs@cmu.edu

**On behalf of 7 correspondents (see go.nature.com/n6gont for full list).*

NEURODEGENERATION

Problems at the nuclear pore

Expansion of a repetitive DNA sequence is associated with neurodegeneration. Three studies identify genes involved in nuclear import and export that can mediate the toxicity this expansion causes. [SEE ARTICLE P.56](#) & [LETTER P.129](#)

BENNETT W. FOX & RANDAL S. TIBBETTS

Advances in molecular genetics and stem-cell technology are transforming our understanding of disease. Such progress is desperately needed in amyotrophic lateral sclerosis, a paralysing neurodegenerative disease that is almost uniformly fatal. It is therefore welcome news that three studies (two in this issue^{1,2} and one in *Nature Neuroscience*³) have converged on a molecular mechanism that seems to underlie a familial form of the disease.

Amyotrophic lateral sclerosis (ALS) is typically sporadic, but around 10% of cases are familial. Although mutations in more than a dozen genes can be involved, those in *C9ORF72* are by far the most common cause of ALS, being responsible for approximately 40% of familial cases⁴. Mutations in *C9ORF72* occur in a section of DNA comprising six tandemly repeated bases: four guanines (G) followed by two cytosines (C). This G₄C₂ hexanucleotide sequence is typically repeated two or three times, but can be expanded to tens or even thousands of repeats in people with *C9ORF72*-associated ALS (C9-ALS)^{5,6}.

There are two leading models proposing how G₄C₂-hexanucleotide-repeat expansion (HRE) leads to neurodegeneration. One, the toxic RNA model, posits that G₄C₂ RNAs transcribed from the expansion bind to crucial RNA-binding proteins or other cellular factors, which prevents them from functioning normally. The other model proposes that, through an unusual form of translation, these expanded RNA molecules produce toxic dipeptide repeat proteins (DPRs) — such as strings of glycine and arginine (GR) amino acids, or of proline and arginine (PR). Central to both hypotheses is the fact that RNAs harbouring the HRE assemble into G-quadruplex structures that confer abnormal molecular behaviours⁷. There is experimental evidence to support both hypotheses^{8–17}, although it is unclear whether DPRs are expressed at sufficient levels to contribute to toxicity¹⁸.

In the first of the three studies, Zhang *et al.*¹ (page 56) engineered *Drosophila melanogaster* fruit flies to express HREs of 30 repeats (termed (G₄C₂)₃₀) in the flies' eyes. G₄C₂ expansion in this system causes neurodegeneration

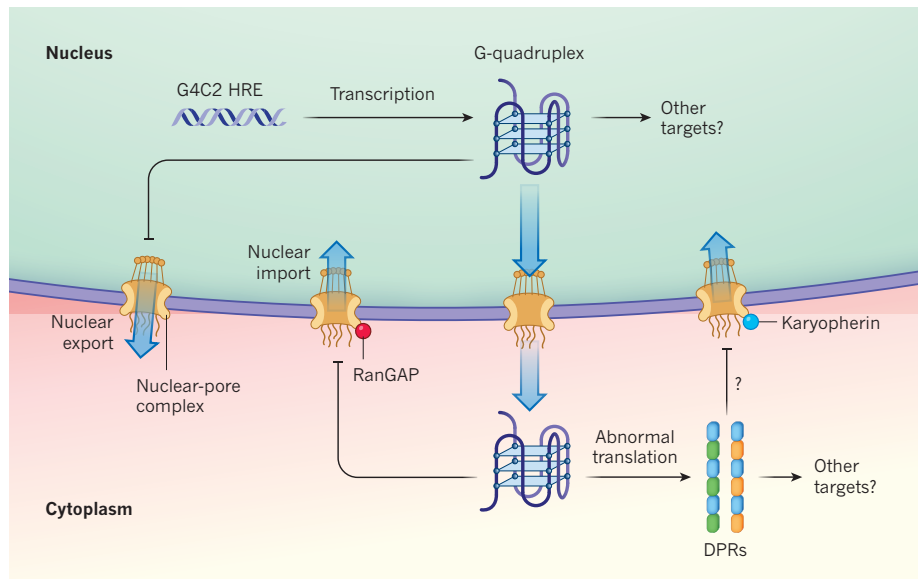


Figure 1 | Models of G₄C₂-mediated neurotoxicity. The neurodegenerative disease amyotrophic lateral sclerosis can be caused by the expansion of a repetitive six-base DNA sequence (GGGGCC; G₄C₂) in the gene *C9ORF72*. Transcription of this G₄C₂ hexanucleotide repeat expansion (HRE) produces RNA that assembles into harmful G-quadruplex structures that can remain in the nucleus or be exported to the cytoplasm. Three studies^{1–3} provide evidence suggesting that the G₄C₂ HRE RNA may cause cell toxicity in several ways: by binding to target proteins in nuclear-pore complexes in the nucleus, thereby disrupting nuclear export of RNA; by binding to the nuclear-import protein RanGAP outside the nucleus, preventing normal nuclear import; and by undergoing an abnormal form of translation to form toxic dipeptide repeat proteins (DPRs) that interfere with nuclear import, possibly through interactions with karyopherin proteins. Other effects of G₄C₂ HRE at both the RNA and protein levels remain unclear.

and results in a 'rough-eye' trait that can be scored to identify mutations in other genes that lessen or worsen toxicity. Reasoning that proteins that bind to the G₄C₂ RNA are probably mediators of G₄C₂-associated toxicity, the authors crossed flies expressing (G₄C₂)₃₀ with flies harbouring mutations in genes encoding G₄C₂-binding proteins identified through previous biochemical screens^{9,10}. Mutations that activated a gene called *RanGAP* strongly suppressed rough eye and neurodegeneration in (G₄C₂)₃₀ flies. The RanGAP protein, which binds to G₄C₂ RNA, is located on the cytoplasmic face of the nuclear membrane, and is part of one of around 2,000 nuclear-pore complexes that control the flow of proteins and RNA in and out of the nucleus. Zhang and colleagues then identified several other nuclear-import genes involved in (G₄C₂)₃₀-elicited neurotoxicity.

Freibaum *et al.*² (page 129) reached similar conclusions using a different strategy. They crossed flies expressing a (G₄C₂)₅₈ HRE with flies missing defined chromosomal segments spanning from tens of genes to more than a hundred, and, using an iterative approach, homed in on genes whose deletion altered G₄C₂ toxicity. The authors identified numerous nuclear-import factors whose inactivation worsened the rough-eye trait of (G₄C₂)₅₈ flies. The comprehensive screen revealed that changes to nuclear-export proteins also enhanced neurodegeneration and rough eye, suggesting that alterations in both nuclear import and export contribute to G₄C₂-mediated toxicity, at least in their system.

Both groups then probed exactly how G₄C₂ HREs disrupt nuclear trafficking. Zhang *et al.* showed that RanGAP bound G₄C₂ HREs *in vitro*, and corroborated this

finding *in vivo* using neurons taken from patients with C9-ALS. HRE expression disrupted the nuclear import of fluorescent test substrates and of normal nuclear proteins — most notably TDP-43, which forms misfolded aggregates in the degenerating motor neurons of most people with ALS. Freibaum and colleagues observed nuclear-membrane irregularities in HRE-expressing cells and demonstrated that G_4C_2 HREs inhibit nuclear RNA export, an effect that was relieved by reducing the expression of genes that suppressed G_4C_2 -mediated toxicity. Together, these findings established a strong connection between defective nuclear trafficking and neurodegeneration (Fig. 1).

Do DPRs contribute to G_4C_2 toxicity? Both studies detected G_4C_2 -derived DPRs, but neither could show whether these DPRs contributed significantly to toxicity. In the third study discussed here, Jovičić *et al.*³ addressed this point directly, performing a genetic screen to identify genes that lessened or worsened the toxicity caused by a PR₅₀ DPR in the yeast *Saccharomyces cerevisiae*. Because PR₅₀ was expressed from synthetic DNA and not from a G_4C_2 HRE, toxicity should derive from the DPR itself, rather than from its parent RNA. Six of the strongest suppressors of PR₅₀-associated toxicity in the researchers' screen encoded members of the karyopherin family of nuclear-import proteins. The screen also suggested that the genesis of ribosomes (the cellular machinery that produces proteins) goes awry in PR₅₀-expressing yeast. In the future, even more leads are likely to be mined from these genetic data.

These three studies take us to a higher plane of understanding of C9ORF72-associated ALS, with a focus placed squarely on the nuclear pore. For the future, the newly identified toxicity-suppressing genes will need to be tested in mammalian models of G_4C_2 expansion and DPR toxicity, probably using recently developed mouse strains¹⁹. The findings also raise the question of whether nuclear-trafficking defects contribute to neurotoxicity in other types of ALS. Neurons have a limited ability to replace damaged nuclear-pore complexes, and age-dependent decreases in nuclear integrity have been postulated as a risk factor for ageing-related disease²⁰. Thus, enhancers of nuclear import should be tested in other ALS models, particularly those in which TDP-43 aggregation is observed.

The genetic studies have not resolved whether one mechanism of toxicity predominates in C9-ALS. At face value, the data suggest that G_4C_2 -containing RNAs and G_4C_2 -derived DPRs elicit toxicity through an overlapping set of nuclear-pore proteins. However, it remains possible that DPRs contribute to neurotoxicity directly in flies. This question could be answered by investigating whether the nuclear-import enhancers picked up in the G_4C_2 screens can rescue neurodegeneration in flies

expressing toxic DPRs. It will also be important to further characterize G_4C_2 -RanGAP interactions, and to determine whether DPRs bind nuclear-pore proteins. Finally, because both DPRs and G_4C_2 HREs reportedly disrupt a subnuclear structure called the nucleolus^{10,13}, the relationship between this mechanism and nuclear-membrane defects should be deciphered.

Can our understanding of toxic G_4C_2 RNA be leveraged for therapy? Zhang *et al.* reversed the rough-eye trait by feeding (G_4C_2)₃₀ flies either a compound that disrupts G_4C_2 -RanGAP binding or a small-molecule inhibitor of nuclear export. The three studies also identified other genes that may be 'druggable', including those encoding proteins that oppose RanGAP. Development and preclinical testing of modulators of nuclear import or export is certainly warranted. No doubt, genetic studies such as the three discussed here will identify other nodes of therapeutic interest. ■

Bennett W. Fox and Randal S. Tibbetts
are in the Department of Human Oncology,
University of Wisconsin–Madison,

Madison, Wisconsin 53706, USA.
e-mail: rstibbetts@facstaff.wisc.edu

1. Zhang, K. *et al.* *Nature* **525**, 56–61 (2015).
2. Freibaum, B. D. *et al.* *Nature* **525**, 129–133 (2015).
3. Jovičić, A. *et al.* *Nature Neurosci.* <http://dx.doi.org/10.1038/nn.4085> (2015).
4. Renton, A. E., Chio, A. & Traynor, B. J. *Nature Neurosci.* **17**, 17–23 (2014).
5. Renton, A. E. *et al.* *Neuron* **72**, 257–268 (2011).
6. DeJesus-Hernandez, M. *et al.* *Neuron* **72**, 245–256 (2011).
7. Simone, R., Fratta, P., Neidle, S., Parkinson, G. N. & Isaacs, A. M. *FEBS Lett.* **589**, 1653–1668 (2015).
8. Zu, T. *et al.* *Proc. Natl Acad. Sci. USA* **110**, E4968–E4977 (2013).
9. Donnelly, C. J. *et al.* *Neuron* **80**, 415–428 (2013).
10. Haeusler, A. R. *et al.* *Nature* **507**, 195–200 (2014).
11. Mori, K. *et al.* *Science* **339**, 1335–1338 (2013).
12. Ash, P. E. *et al.* *Neuron* **77**, 639–646 (2013).
13. Kwon, I. *et al.* *Science* **345**, 1139–1145 (2014).
14. Mizielinska, S. *et al.* *Science* **345**, 1192–1194 (2014).
15. Lagier-Tourenne, C. *et al.* *Proc. Natl Acad. Sci. USA* **110**, E4530–E4539 (2013).
16. Sareen, D. *et al.* *Sci. Transl. Med.* **5**, 208ra149 (2013).
17. Su, Z. *et al.* *Neuron* **83**, 1043–1050 (2014).
18. Gomez-Deza, J. *et al.* *Acta Neuropathol. Commun.* **3**, 38 (2015).
19. Chew, J. *et al.* *Science* **348**, 1151–1154 (2015).
20. D'Angelo, M. A., Raices, M., Panowski, S. H. & Hetzer, M. W. *Cell* **136**, 284–295 (2009).

This article was published online on 26 August 2015.

SOFT MATTER

Frictionless fluids from bacterial teamwork

By increasing the sensitivity of an established technique, researchers have shown that swimming bacteria can make frictionless fluids — with potential applications in areas such as microfluidics.

M. CRISTINA MARCHETTI

The viscosity of a liquid is a measure of its resistance to flow. In general, denser fluids are more viscous and require more energy to get them to flow through a pipe. Flow with no energy dissipation is a hallmark of exotic states of matter such as superfluidity and superconductivity. Key to these exotic states are quantum effects that dominate at ultralow temperatures — turning liquid helium, for example, into a superfluid that flows without friction through cracks as thin as molecules. Writing in *Physical Review Letters*, López *et al.*¹ demonstrate that *Escherichia coli* bacteria swimming in a fluid can organize themselves to counterbalance the energy loss resulting from viscous dissipation and thereby dramatically lower the fluid's viscosity, driving it to vanish or even to become negative.

In 2004, it was predicted² that unicellular swimming organisms could substantially

change the viscosity of a fluid on the basis of a hydrodynamic theory³ of active fluids (liquids consisting of self-propelled particles). This suggestion was confirmed by numerical solutions^{4,5} of the theory, which revealed the possibility of vanishing viscosity for suspensions of motile bacteria. Pioneering experiments subsequently confirmed a reduction of viscosity in suspensions of the bacteria *Bacillus subtilis*⁶ and *E. coli*⁷. A concurrent study⁸ demonstrated the sensitivity of this effect to the microscopic cellular-propulsion mechanism by revealing an increase in viscosity for dilute concentrations of the alga *Chlamydomonas reinhardtii*; however, this behaviour remains puzzling.

Detailed calculations^{9–12} of the response of dilute suspensions of swimmers to an externally imposed shear flow (which induces the velocity profile shown in Fig. 1a) have provided quantitative expressions for viscosity changes for small volume fractions of bacteria. Demonstrating that a bacterial suspension can

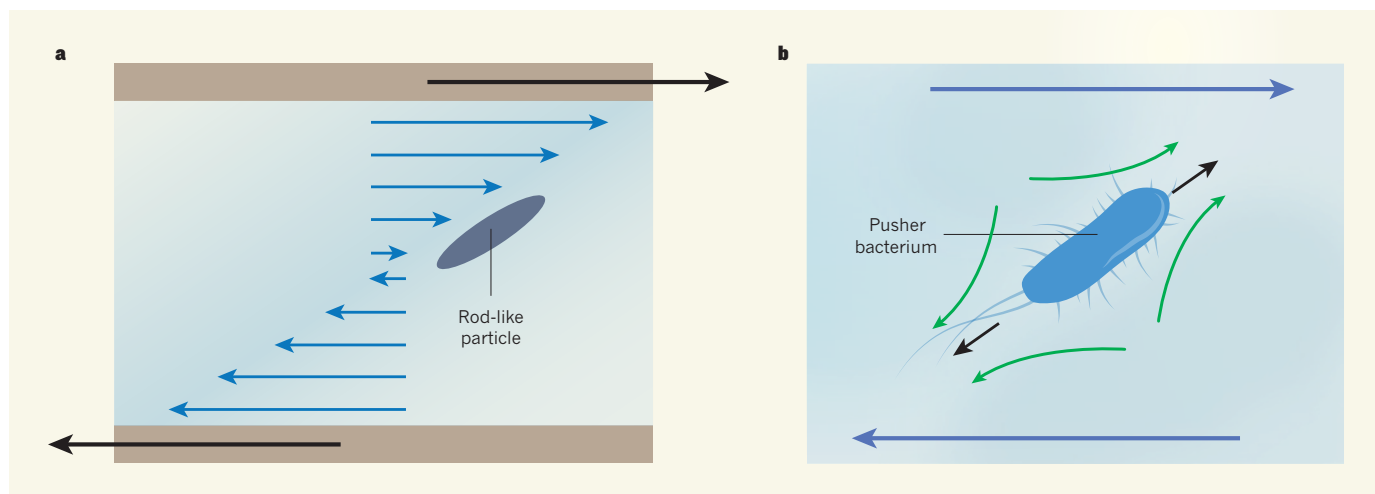


Figure 1 | Viscosity modulation by rod-like bacteria. **a**, Shear flow (blue arrows) can be imposed on a fluid between two plates by applying equal and opposite forces (black arrows) on the plates. Rod-like particles suspended in the liquid respond by orienting their long axes along the direction in which the flow tends to 'stretch' the fluid. **b**, Bacteria such as *Escherichia coli* are 'pusher' swimmers, which use their rear flagella appendages for propulsion.

Pushers exert equal and opposite forces (black arrows) at their front and tail. In an external shear flow (blue arrows), they orient along the stretching direction of the flow and generate flow fields (green arrows) that further push the fluid in the same direction as the shear flow, reducing the fluid's viscosity. López *et al.*¹ report that suspensions of *E. coli* can generate fluids with zero, or even negative, viscosity.

achieve a state of vanishing, or even negative, viscosity was previously impossible, however, because this requires measurements of tiny shear stresses.

López and colleagues overcame this problem by adapting an old-fashioned rheometer — a device used to measure fluid viscosity. A simple rheometer consists of inner and outer cylinders that can rotate relative to each other. A fluid is placed in the annulus between the cylinders and one of the cylinders is rotated at a set rate, shearing the fluid. The liquid drags the other cylinder, exerting a torque on it. From measurements of this torque, one can infer the shear stress and thus the fluid viscosity, defined as the ratio of stress to the applied shear rate.

The authors modified this device by controlling the rotation of the inner cylinder using a computerized feedback mechanism. This set-up maintains zero torque, allowing highly sensitive measurements of ultralow shear stresses. The researchers also suspended bacteria in a medium that allows the microbes to remain motile but not to divide, enabling control of the bacterial concentration.

Importantly, López *et al.* were able to demonstrate the existence of states of arbitrarily small viscosity, in a regime in which the viscosity did not depend on the imposed shear rate and was therefore a legitimate material property. The development of a macroscopic device capable of sensing the rheological response of microorganisms is a remarkable experimental achievement. It paves the way for the quantitative characterization of the flow behaviour of a wide range of microorganisms, and for understanding the role of different propulsion mechanisms.

Suspending non-motile particles in a fluid always increases the fluid's viscosity.

Albert Einstein provided the first quantitative formulation of this intuitive effect in 1906 by showing that, in a dilute suspension of spheres, the increase in viscosity is linearly proportional to the volume fraction of suspended particles¹³. So how do swimming bacteria achieve the opposite effect and thin out the suspension, turning it into a frictionless liquid akin to a superfluid?

The answer relies on two key properties of flowing suspensions. First, inactive rod-like particles in an externally imposed shear flow align their long axes along the direction in which the flow 'stretches' the fluid. The rods tilt at a fixed angle that depends on their ratio of length to width; this angle can be close to 45° for long, slender rods (Fig. 1a). Many unicellular organisms, including *E. coli*, have such a rod-like shape and therefore orient in this way in shear flow.

Second, swimming bacteria exert forces on the surrounding fluids. These forces come in equal and opposite pairs: the force from the beating of their propulsive appendages (flagella or cilia) is balanced by the viscous drag on the cell's body. The spatial profile of these forces depends on the propulsion mechanism. Most bacteria use appendages mounted at the back of their bodies, and are known as pushers. When they move, they push fluid out at their front and back, while sucking it in at the sides.

Elongated pushers thus align their bodies along the stretching axis of the external flow and generate additional flows that further stretch the fluid in the same direction (Fig. 1b). At high enough concentrations, continuum theories suggest that the bacteria act collectively to push the fluid along, effectively thinning it. A microscopic understanding of how bacteria coordinate their response to

shear to achieve a state of frictionless flow is still lacking. However, López and co-workers' experiments demonstrate that the viscosity of a suspension of swimming bacteria can indeed decrease with an increasing volume fraction of swimmers, within a range of bacterial concentrations.

By showing that bacteria can completely compensate for fluid friction by allowing the fluid to flow with zero dissipation, the authors have demonstrated that it is possible, in principle, to extract useful macroscopic mechanical power from bacterial activity. This observation is in line with earlier findings¹⁴ that bacteria can work together to turn microgears. Although harnessing bacterial power for macroscopic energy generation may still be a dream, it is not such a stretch to imagine that bacteria could be used as mixers to thin and stir the flow in capillary and microfluidic devices. Quantitative characterizations of rheology of the type pioneered by López *et al.* pave the way to the development of bacterial baths tailored to mix and flow liquids for specific applications. ■

M. Cristina Marchetti is in the Department of Physics and Syracuse Biomaterials Institute, Syracuse University, Syracuse, New York 13244, USA.
e-mail: mcmarche@syr.edu

1. López, H. M., Gachelin, J., Douarche, C., Auradou, H. & Clément, E. *Phys. Rev. Lett.* **115**, 028301 (2015).
2. Hatwalne, Y., Ramaswamy, S., Rao, M. & Simha, R. A. *Phys. Rev. Lett.* **92**, 118101 (2004).
3. Simha, R. A. & Ramaswamy, S. *Phys. Rev. Lett.* **89**, 058101 (2002).
4. Cates, M. E., Fielding, S. M., Marenduzzo, D., Orlandini, E. & Yeomans, J. *Phys. Rev. Lett.* **101**, 068102 (2008).
5. Giomi, L., Liverpool, T. B. & Marchetti, M. C. *Phys. Rev. E* **81**, 051908 (2010).

6. Sokolov, A. & Aranson, I. S. *Phys. Rev. Lett.* **103**, 148101 (2009).
7. Gachelin, J. et al. *Phys. Rev. Lett.* **110**, 268103 (2013).
8. Rafai, S., Jibuti, L. & Peyla, P. *Phys. Rev. Lett.* **104**, 098102 (2010).
9. Liverpool, T. B. & Marchetti, M. C. *Phys. Rev. Lett.* **97**, 268101 (2006).
10. Haines, B. M., Sokolov, A., Aranson, I. S., Beryland, L. & Karpeev, D. A. *Phys. Rev. E* **80**, 041922 (2009).
11. Saintillan, D. *Exp. Mech.* **50**, 1275–1281 (2010).
12. Ryan, S. D., Haines, B. M., Beryland, L., Ziebert, F. & Aranson, I. S. *Phys. Rev. E* **83**, 050904(R) (2011).
13. Einstein, A. *Ann. Phys.* **19**, 289–306 (1906).
14. Sokolov, A., Apodaca, M. M., Grzybowski, B. A. & Aranson, I. S. *Proc. Natl Acad. Sci. USA* **107**, 969–974 (2009).

ECOLOGY

Global trends in plant naturalization

Many naturalized non-native plants pose ecological and economic threats. A quantitative analysis of the global distribution of naturalized plants confirms some anticipated trends and exposes new patterns. SEE LETTER P.100

MARCEL REJMANEK

Naturalized species are non-native species that form self-sustaining populations following their introduction into an area by human agency¹. Some naturalized species are considered a major threat to biodiversity and have been the focus of many biologists over the past three decades. However, even casual observers may notice that the distribution of naturalized species is highly uneven within and among different regions. Attempts to summarize global geographical distributions of naturalized organisms have included birds², ungulates³ (large mammals such as pigs and camels) and bryophytes⁴ (non-vascular plants), but a comprehensive assessment of naturalized vascular plants has been missing. In this issue, van Kleunen *et al.*⁵ (page 100) provide the first global analysis of the numbers and distributions of naturalized vascular plants and their exchange between continents.

The authors used hundreds of data sources of various kinds to characterize the alien floras of 843 non-overlapping regions worldwide (481 mainland and 362 island areas). Characterization included the origin of the naturalized species and estimates of the numbers of native and non-native species per continent. The resulting database includes 13,168 plant species — 3.9% of the world's currently known vascular flora — that have become naturalized in at least one region. The authors suggest that this figure may be an underestimate, given the lack of data (or adequate data) for some regions.

One of the most striking results of this study comes from the authors' comparisons between large continental areas. These revealed that North America has accumulated the largest number of naturalized species of vascular plant (5,958), followed by Europe (4,140). This finding undoubtedly reflects more intensive

introduction processes — both deliberate, for example for ornamental horticulture and erosion control, and accidental, as a result of frequent trade between these regions and the rest of the world.

Simple numbers of naturalized species do not, however, quantify the actual level of invasion. Previous work⁶ has shown that, in North America, non-native species account for 51.3% of the 120 most widely distributed plant species, but account for only 2.1% in Europe. One possible explanation for the striking difference between Europe and North America is that the European flora, being part of the Eurasian flora, has been exposed to countless plant migrations over time, so that the resulting plant communities are less 'naïve' and more resistant to new plant invasions. It is also likely that some European plant species have been selected for quick colonization of human-disturbed habitats, the habitats in which they are most often found naturalized in North America (Fig. 1).

Van Kleunen and colleagues' data also show that the Pacific Islands region exhibits the steepest increase in the cumulative number of naturalized species with respect to the total area involved. This result provides the first global verification of an expected pattern: that oceanic islands harbour more naturalized plant species than mainland areas of similar size. A primary reason for this may be that native communities on islands represent only a limited sample of the species that could potentially match the habitat, and they are therefore more open to the naturalization of introduced species.

At the same time, the data confirm previous preliminary analyses showing that continental regions with large tropical areas (Africa, South America, tropical Asia) have fewer naturalized plant species than predominantly temperate regions. Higher resistance to non-native species establishment, faster vegetation



50 Years Ago

It is probable that only those who have themselves been concerned with scientific research will appreciate all the fine nuances of Sir Cyril [Hinshelwood]'s address, but the picture he paints of the scientist as a creative worker, of the need for freedom of expression and appropriate conditions of work, and of public understanding if his work is to be fully effective, is intelligible to any layman. It is no picture of a scientist working and living in some 'ivory tower', or even of Thomson's Newton, "stemming alone vast eternity's unbounded sea", but rather of a happy voyager of strange seas of thought, in company with others trained in the same or many other disciplines.

From *Nature* 4 September 1965

100 Years Ago

In his presidential address, read at the Association of Museums, San Francisco, Dr. O. C. Farrington gave an able summary of the origin and evolution of natural history museums, which should be widely read in this country. More especially is this to be urged in view of the danger which threatens such institutions in the immediate future in regard to the policy of national retrenchment which is now in process of formation. There is a danger that the pruning-hook may be used too ruthlessly, thereby inflicting material harm. For reformers are generally enthusiasts, and therefore are to be carefully watched, experience having shown that a sense of proportion is not usually among their attributes. Museums, as he remarks, are even now commonly regarded as a luxury, but he leaves no uncertainty as to the vitally important part which the modern museum plays, and must continue to play, in ever-increasing force, in our national life.

From *Nature* 2 September 1915



MARCEL REJMANEK

Figure 1 | A naturalized problem. *Hypericum canariense* L., a species of St John's wort, is native to the Canary Islands, but has been introduced to and become naturalized in California (left). Almost no plants can grow beneath stands of this naturalized shrub (right). Van Kleunen and colleagues' analysis⁵

of naturalized plants across the globe reveal that California is a region with one of the highest numbers of naturalized species. Most of them are incorporated into native, mostly highly human-disturbed, biotic communities without any obvious impacts. However, some, like *H. canariense*, are highly influential.

recovery following disturbance and lower introduction intensity in the tropical regions are the most common hypotheses explaining this phenomenon.

Another finding of van Kleunen and colleagues' study is that the continents of the Northern Hemisphere have been the major sources of naturalized plant species for many areas of the Southern Hemisphere, but not vice versa. Thus, the study finally quantifies what Charles Darwin anticipated on the basis of the observations he made on his voyage around the world. Darwin proposed⁷: "I suspect that this preponderant migration from north to south is due to the greater extent of land in the north, and to the northern forms having existed in their own homes in greater numbers, and having consequently been advanced through natural selection and competition to a higher stage of perfection or dominating power, than the southern forms." But conclusively testing this hypothesis will be difficult, if not impossible — hundreds, at least, of phylogenetically related pairs of species from the Northern and Southern Hemispheres would have to be tested in well-designed competition experiments. Alternatively, the fact that the Southern Hemisphere is currently underrepresented as a source of naturalized vascular plants may indicate that southern continents provide many species that could spread to the northern continents in the future.

The quality of the data used for this synthesis could be much improved — for example, the definition of naturalized species is not always clear and consistent across different sources of data; not all sources are equally reliable; and the botany of some regions is much less well known than others. Nevertheless, it seems that van Kleunen and colleagues' major qualitative conclusions are robust. Two straightforward generalizations emerge from this and other studies: with an increasing

number of introduced species, the number of naturalized species increases⁸; and with an increasing number of naturalized species, the number of potentially harmful species also increases⁹. Moreover, the number of naturalized plant species and their combined cover¹⁰ are usually positively correlated (at least in small plots in North America). It is probable that the data collected in this study will be used for testing many interesting hypotheses¹¹ and will improve our predictions of the future distributions of naturalized plant species¹². ■

Marcel Rejmánek is in the Department of Evolution and Ecology, University of California Davis, Davis, California 95616, USA.

e-mail: mrejmanek@ucdavis.edu

1. Richardson, D. M. *et al. Divers. Distrib.* **6**, 93–107 (2000).
2. Case, T. J. *Biol. Conserv.* **78**, 69–96 (1996).
3. Spear, D. & Chown, S. L. *J. Zool.* **279**, 1–17 (2009).
4. Essl, F., Dullinger, S., Moser, D., Steinbauer, K. & Mang, T. *Ecography* **38**, 488–498 (2015).
5. van Kleunen, M. *et al. Nature* **525**, 100–103 (2015).
6. Stohlgren, T. J. *et al. Biol. Invasions* **13**, 1931–1944 (2011).
7. Darwin, C. *On the Origin of Species* (John Murray, 1859).
8. Hulme, P. E. *J. Appl. Ecol.* **49**, 10–19 (2012).
9. Rejmánek, M. & Randall, J. M. *Divers. Distrib.* **10**, 367–369 (2004).
10. Seabloom, E. W. *et al. Global Change Biol.* **19**, 3677–3687 (2013).
11. Fridley, J. D. & Sax, D. F. *Global Ecol. Biogeogr.* **23**, 1157–1166 (2014).
12. Gallien, L., Münckmüller, T., Albert, C. H., Boulangeat, I. & Thuiller, W. *Divers. Distrib.* **16**, 331–342 (2010).

This article was published online on 19 August 2015.

SUPERCONDUCTIVITY

Extraordinarily conventional

Attitudes to high-temperature superconductivity have swung from disbelief to a conviction that it occurs only 'unconventionally'. But conventional superconductivity is now reported at record high temperatures. [SEE LETTER P.73](#)

IGOR I. MAZIN

In 1911, the physicist Heike Kamerlingh Onnes was puzzled to observe¹ that mercury became an ideal conductor below 4.2 kelvin. How could all the electrons in a metal cooperate so as to carry electric current without resistance? Common wisdom

dictates that there is nothing ideal in this world. Nobody's perfect! No crystals without defects can be created, no wheel can roll without friction, no glass can be 100% transparent. Yet subsequent experiments confirmed that the resistivity of many metals suddenly drops to exactly zero at a sufficiently low temperature. Chaotic motion introduced by

heat destroys electronic cooperation, so for many years it was believed that this phenomenon, now known as superconductivity, is limited to ultra-low temperatures. But on page 73 of this issue, Drozdov and colleagues² report a superconductor that works at about 200 K — a temperature that actually exists on Earth's surface.

For decades, physicists were in the dark about the origin of superconductivity. The discovery in 1938 of another cooperative phenomenon³ — superfluidity in helium at 2 K — offered the first clue⁴. This complete lack of viscosity turned out to be a direct consequence of quantum mechanics. All quantum particles are characterized by a 'spin' number; if this is an integer (as for helium-4 atoms), the particles can combine into a single object so large that it cannot be disturbed by such nuisances as friction or viscosity. This effect is called Bose–Einstein condensation.

But electrons, which conduct electricity, have a spin of $\pm\frac{1}{2}$, and so are not subject to Bose–Einstein condensation. In 1957, John Bardeen, Leon Cooper and Bob Schrieffer therefore proposed that the interaction of electrons with metal ions creates an attractive interaction that forces the electrons to combine in pairs⁵. These 'Cooper pairs' have a net spin of zero, and can form a Bose–Einstein condensate. This theory also allows the transition temperature, T_c , below which superconductivity occurs for a given metal, to be estimated.

Most elemental superconductors had been discovered by 1957, and all had T_c values of less than 10 K. For the next two decades, scientists worked with various compounds, but failed to increase T_c by even a factor of three. Not surprisingly, most physicists began to believe that nature imposes a fundamental, but as-yet unexplained, T_c limit of 25–30 K. The problem was succinctly formulated by the materials scientist Bernd Matthias in 1964: "Why has it been relatively easy, within the last 10 years, to reach transition temperatures of 17 to 18 K in many intermetallic systems and impossible to raise this value even by as little as half a degree?"⁶ Eight years later, Marvin Cohen and Phillip Anderson pointed out that if electrons interact too strongly with the ions in a metal, they can break the lattice apart⁷. On this basis, they estimated that the highest T_c for conventional superconductors (those driven by the electron–ion interaction) is approximately 30 K.

Although the argument seemed convincing, some physicists remained hesitant. In the early 1970s, Vitaly Ginzburg — one of the top theorists of the time — organized a group in Moscow to explore routes to high-temperature superconductivity. One of his team's principal results was that a key assumption by Cohen and Anderson was flawed, and that T_c could, in principle, be arbitrarily high even in a conventional superconductor⁸.

Another prominent physicist who did not subscribe to the idea of a universal limit was Neil Ashcroft. In the late 1960s, he⁹ and Ginzburg¹⁰ proposed that, if hydrogen could become metallic, the energy of its ionic vibrations would be so high that even a moderately strong electron–ion coupling could result in a rather high T_c . Unfortunately, metallization of hydrogen has proved to be extremely difficult. It was then pointed out that hydrogen-rich compounds might be better targets^{11,12}, but it is only now that this idea has been realized, as reported by Drozdov and colleagues.

In the meantime, three major breakthroughs occurred in superconductivity. First, cuprate superconductors were discovered¹³ in 1986; within seven years, the T_c for these compounds reached 133 K (ref. 14). These have been recognized as 'unconventional' superconductors, driven by interactions among electrons, rather than by electron–ion interactions.

The second was the discovery¹⁵, in 2001, of magnesium diboride — a conventional superconductor whose T_c is 40 K. This relatively high number is due to the low mass of boron, and to the fact that strong electron–ion coupling

is ensured because the conducting electrons come from the boron, and the boron ions form a rigid sublattice. The physics of superconductivity in magnesium diboride turned out to be considerably more complex than

for other conventional superconductors known at the time, but was understood within a year of the original discovery. At last, theorists could accurately calculate the critical temperature of a rather complicated material. This encouraged scientists to seek quantitative predictions for new superconducting materials.

The third breakthrough was the discovery¹⁶ of iron-based superconductors in 2008. These materials seem to be unconventional and, although of great interest, have never surpassed the T_c of the cuprates.

Drozdov and co-workers report a fourth breakthrough: superconductivity at approximately 200 K in a hydrogen-rich compound, sulfur hydride, at about 90 gigapascals — a pressure hardly achievable just a few years ago. Not only is this a 50% increase over the previous record for T_c , but the authors convincingly argue that the observed superconductivity is conventional, vindicating the ideas of Ashcroft and Ginzburg.

Moreover, this is the first time that a previously unknown material predicted to be a high-temperature superconductor has been experimentally confirmed to be one. A computational study¹⁷ of hydrogen-rich materials under pressure had reported that the sulfur

hydride H_3S would be a superconductor with T_c in the range 190–200 K at 200 gigapascals — very close to the now-reported experimental value. Drozdov *et al.* studied H_2S , but it seems that at high pressure this decomposes into elemental sulfur and hydrogen-rich H_3S . It is therefore highly likely that the superconducting material is H_3S . More-accurate calculations^{18–20} yielded a T_c value approximately 20% higher than in the earlier computational study¹⁷. There is some disagreement about which small effects, not accounted for in standard computations, are responsible for this overestimate, but it is amazing that theorists quibble about a 20% inaccuracy in first-principles calculations when even an order-of-magnitude estimation was considered practically impossible only 40 years ago.

In 1796, the philosopher Wilhelm Hegel introduced the concept of spiral progress: an intellectual proposition is superseded by its negation, but later the negation itself is negated; the original thesis is then reinstated, but at a higher level of development. The generality of this concept can be philosophically disputed, but Hegel's idea seems to be confirmed by the fact that the holy grail of superconductors has been discovered in the same group of materials as the first known superconductors, after a tiresome quest along exotic routes. ■

Igor I. Mazin is at the Center for Computational Materials Science, Naval Research Laboratory, Washington, DC 20375-5320, USA.
e-mail: mazin@dave.nrl.navy.mil

1. Kamerlingh Onnes, H. *Commun. Phys. Lab. Univ. Leiden* No. 120 (1911).
2. Drozdov, A. P., Erements, M. I., Troyan, I. A., Ksenofontov, V. & Shylin, S. I. *Nature* **525**, 73–76 (2015).
3. Kapitza, P. *Nature* **141**, 74 (1938).
4. London, F. *Nature* **141**, 643–644 (1938).
5. Bardeen, J., Cooper, L. N. & Schrieffer, J. R. *Phys. Rev.* **108**, 1175 (1957).
6. Anderson, P. W. & Matthias, B. T. *Science* **144**, 373–381 (1964).
7. Cohen, M. L. & Anderson, P. W. in *Superconductivity in d- and f-Band Metals* (ed. Douglass, D. H.) 17–27 (American Inst. Physics, 1972).
8. Dolgov, O. V., Kirzhnits, D. A. & Maksimov, E. G. *Rev. Mod. Phys.* **53**, 81 (1981).
9. Ashcroft, N. W. *Phys. Rev. Lett.* **21**, 1748 (1968).
10. Ginzburg, V. L. *J. Stat. Phys.* **1**, 3–24 (1969).
11. Ginzburg, V. L. & Kirzhnits, D. A. (eds) *High-Temperature Superconductivity* Chapter 1 (Consultants Bureau, 1982) [transl. from Russian].
12. Ashcroft, N. W. *Phys. Rev. Lett.* **92**, 187002 (2004).
13. Bednorz, J. G. & Müller, K. A. *Z. Physik B* **64**, 189–193 (1986).
14. Schilling, A., Cantoni, M., Guo, J. D. & Ott, H. R. *Nature* **363**, 56–58 (1993).
15. Nagamatsu, J., Nakagawa, N., Muranaka, T., Zenitani, Y. & Akimitsu, J. *Nature* **410**, 63–64 (2001).
16. Kamihara, K. Y., Watanabe, T., Hirano, M. & Hosono, H. *J. Am. Chem. Soc.* **130**, 3296–3297 (2008).
17. Duan, D. *et al. Sci. Rep.* **4**, 6968 (2014).
18. Errea, I. *et al. Phys. Rev. Lett.* **114**, 157004 (2015).
19. Flores-Livas, J. A., Sanna, A. & Gross, E. K. U. Preprint at <http://arxiv.org/abs/1501.06336> (2015).
20. Akashi, R., Kawamura, M., Tsuneyuki, S., Nomura, Y. & Arita, R. *Phys. Rev. B* **91**, 224513 (2015).

This article was published online on 17 August 2015.

GENETICS

Location affects sporulation

Monitored changes in the number of copies of a gene during DNA replication control the timing of sporulation in bacteria. This discovery links replication to the concept that a gene's location on a chromosome can influence cell traits.

BETH A. LAZAZZERA & DIARMAID HUGHES

For decades it has been known¹ that the location of a gene on its chromosome can influence the level at which it is expressed. Most bacterial chromosomes are circular, and their replication begins at a single bidirectional origin. As such, during chromosome replication, genes close to the origin of replication will be transiently present in more copies (present at a higher dosage) than those close to the terminus of replication. Altering the distance of a gene from the origin of replication systematically alters its level of expression during the cell's replication cycle^{2–4}. But until now, the significance of gene location has largely focused on whether highly expressed genes are preferentially located in the origin-proximal half of the chromosome, because this provides the cell with a growth advantage due to a positive gene-dosage effect⁵. Writing in *Cell*, Narula *et al.*⁶ report a new twist on the role of chromosomal location in gene function, in coordinating sporulation with chromosome replication in the bacterium *Bacillus subtilis*.

When starved, *B. subtilis* can initiate a cascade of protein phosphorylation that leads to sporulation, producing a dormant spore that is resistant to starvation conditions and that can eventually resume growth under

favourable conditions. The first components of this phosphorelay are a kinase enzyme called KinA and a response-regulator protein, Spo0F. *In vitro* evidence⁷ has suggested that, although phosphorylation of Spo0F by KinA is necessary for the activation of early sporulation genes, high concentrations of Spo0F can also inhibit the activity of KinA. Narula *et al.* confirmed this result *in vivo*, demonstrating that high levels of Spo0F induce a negative-feedback loop that inhibits the phosphorelay.

The *spo0F* gene is located near the origin of replication, whereas the *kinA* gene is located close to the replication terminus. Narula *et al.* report that the positions of *spo0F* and *kinA* seem to be crucial for their ability to efficiently regulate sporulation. Because of their respective locations, during replication there is a temporary twofold increase in the dosage of *spo0F* relative to *kinA* (Fig. 1a). By using computer simulations and then verifying their models *in vivo*, the authors showed that the transient increase in Spo0F concentration inhibits KinA until replication is completed, leading to pulsing dynamics of sporulation-gene expression during each cell cycle (Fig. 1b). Cells will only sporulate once they cross a threshold level of sporulation-gene expression, which is achieved through a positive-feedback loop that increases levels of KinA concentration — a process

that takes several rounds of cell division⁸.

Narula and colleagues then performed translocation experiments, in which they moved *spo0F* or *kinA* towards the terminus or origin of replication, respectively. These translocations abolished pulsing, confirming that a transient imbalance in the dosage of the two genes is necessary for pulsing of early sporulation-gene expression and for proper coordination of the sporulation program with DNA replication. These data, together with the authors' finding that the relative locations of *kinA* and *spo0F* are similar in 45 other species of sporulating bacteria, show for the first time that the siting of interacting genes at different locations on the chromosome could have evolved as a way of controlling how the gene products function.

Monitoring chromosome replication status is crucial for many species. In the case of *B. subtilis*, initiation of sporulation without complete chromosomes for both the mother cell and the future spore cell would be a waste of resources. It has long been known⁹ that a checkpoint is activated to inhibit sporulation when DNA is damaged or replication is defective. Narula and colleagues have identified a remarkably simple mechanism by which cells can monitor the replication status of the chromosome.

The regulatory mechanism presented in this study deepens our understanding of the potential variety of mechanisms that might regulate changes in cellular traits. But the work also raises several interesting avenues for further investigation. For example, it is unclear whether this particular situation is a biological one-off. It seems more likely that there are other traits, both in *B. subtilis* and in other organisms, that are regulated by temporal variations in gene-product ratios associated with gene location.

It also remains to be seen whether more-complex versions of this mechanism exist, involving more than two genes, and whether

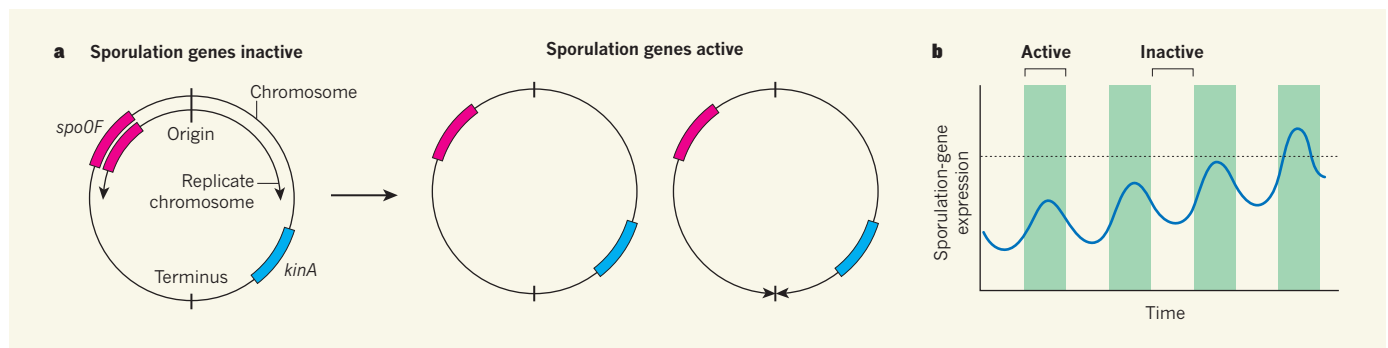


Figure 1 | A genetic imbalance regulates sporulation. The bacterium *Bacillus subtilis* sporulates by activating a phosphorylation cascade, which begins with phosphorylation of the protein Spo0F by the enzyme KinA. **a**, The *spo0F* gene is located close to the origin of DNA replication on the *B. subtilis* chromosome, whereas *kinA* is close to the replication terminus. During replication, the cellular concentration of Spo0F transiently increases relative to the level of KinA, owing to a difference in gene copy number. Narula *et al.*⁶

report that this discrepancy induces a negative-feedback pathway that inactivates sporulation genes. Once replication is complete, the disparity is resolved, and sporulation-gene-expression pathways are activated. **b**, As such, sporulation genes are activated in pulses during sequential cell cycles. Green indicates that the chromosome is fully replicated, background colour indicates partial replication. Once a threshold level of expression is reached (dotted line), sporulation occurs.

such mechanisms could be involved in replication fidelity. For instance, could this type of regulatory mechanism act as a brake on chromosomal rearrangements such as inversions, which might disrupt the relative locations of genes in regulatory networks that rely on dosage imbalances?

Narula and colleagues' work illustrates the potential importance of gene location in perhaps unexpected aspects of cell biology. It will doubtless motivate future experiments in chromosome remodelling. Perhaps it will also prompt a re-examination of old data, to assess whether arbitrary choices made

in genetic engineering might have affected experimental outcomes. ■

Beth A. Lazazzera is in the Department of Microbiology, Immunology, and Molecular Genetics, University of California Los Angeles, Los Angeles, California 90095, USA.

Diarmaid Hughes is in the Department of Medical Biochemistry and Microbiology, Uppsala University, Uppsala 751 23, Sweden. e-mails: bethl@microbio.ucla.edu; diarmaid.hughes@imbim.uu.se

1. Bremer, H. & Dennis, P. P. in *Escherichia coli* and

- Salmonella Vol. 2 (ed. Neidhardt, F. C.) 1553–1569 (ASM Press, 1996).
2. Schmid, M. B. & Roth, J. R. *J. Bacteriol.* **169**, 2872–2875 (1987).
3. Sousa, C., De Lorenzo, V. & Cebollat, A. *Microbiology* **143**, 2071–2078 (1997).
4. Block, D. H. S., Hussein, R., Liang, L. W. & Lim, H. N. *Nucleic Acids Res.* **40**, 8979–8992 (2012).
5. Couturier, E. & Rocha, E. P. C. *Mol. Microbiol.* **59**, 1506–1518 (2006).
6. Narula, J. et al. *Cell* **162**, 328–337 (2015).
7. Grimshaw, C. E. et al. *Biochemistry* **3**, 1365–1375 (1998).
8. Levine, J. H., Fontes, M. E., Dworkin, J. & Elowitz, M. B. *PLoS Biol.* **10**, e1001252 (2012).
9. Veening, J. W., Murray, H. & Errington, J. *Genes Dev.* **23**, 1959–1970 (2009).

This article was published online on 19 August 2015.

HYDROLOGY

The diversified economics of soil water

Soil water that evaporates or is tapped by plants is largely separate from that which runs into streams and recharges groundwater. This finding has big implications for our understanding of water cycling. [SEE LETTER P.91](#)

GABRIEL BOWEN

Soils can be viewed as the investment managers of the terrestrial water cycle: they accept precipitation capital from the atmosphere and allocate it to sustain and grow various biological and hydrological stocks. These water investments influence plant productivity, run-off to streams and groundwater, and atmospheric humidity, so deciphering how soils partition water is vital if we are to understand and predict the function of these systems. On page 91 of this issue, Evaristo *et al.*¹ suggest that the allocation of water in most soils worldwide follows a conservative, diversified 'strategy', in which new resources are invested as they are obtained, and transfer of capital between accounts is limited.

Water researchers have considered two contrasting scenarios for soil-water allocation. The 'commingled' scenario, which is the one most widely adopted in hydrological models (see refs 2 and 3, for example), assumes that all water is held in a common pool and is withdrawn only as needed. Residual water from past precipitation is tapped by plants, evaporates or hosts biogeochemical reactions until fresh precipitation displaces some or all of it into groundwater reservoirs or streams. This situation has been referred to as hydrological connectivity, because water leaving the soil in any form is drawn from a common pool and is connected to all other flows.

The contrasting scenario could be said to be 'diversified', because new water is allocated to

one of several pools as it enters the soil, and transfers between these pools are limited. Previous hydrological research has provided hints of a diversified approach to soil-water investment. For example, in many soils a substantial fraction of infiltrating water moves rapidly through large pores to recharge groundwater and produce stream run-off⁴. There is surprising evidence that this separation of soil-water pools extends to water withdrawal by plants, with trees and shrubs in two ecological systems^{5,6} drawing from a soil-water

pool that is apparently distinct from that feeding recharge and run-off. The generality and importance of such diversification have been unclear, however, because relatively few studies have been conducted.

Evaristo and colleagues provide compelling evidence that the diversified mode is widespread, if not ubiquitous. The authors adapt previously reported methods^{5,6} that capitalize on a distinctive shift in the ratios of hydrogen isotopes and oxygen isotopes in soil water as it evaporates from soils. If the water in soil, plants, groundwater and streams all showed a common evaporation shift, this would strongly suggest a commingled situation (Fig. 1a). But in a meta-analysis of data from 47 studies spanning multiple environments and biomes, the authors instead find a similar evaporation shift in soil and plant water, but little or no shift in streams and groundwater (Fig. 1b). This implies diversification: plants and soil evaporation across ecosystems seem to be tapping a pool of water that is largely separate from the pools that generate run-off and recharge.

This finding has major implications

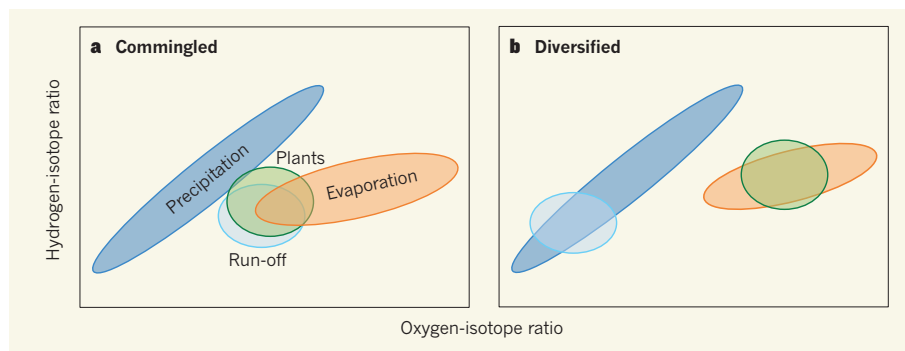


Figure 1 | Evaporation shifts in two models of soil-water allocation. When the ratios of the abundances of hydrogen and oxygen isotopes in soil water are plotted against each other, different data distribution patterns are expected, depending on how the water is partitioned for use. **a**, In the 'commingled' scenario, soil water is held in a common pool. The distributions associated with water used for different purposes — water tapped by plants, run-off to streams and groundwater, or evaporated water — are shifted by similar amounts from the distribution associated with precipitation. **b**, In the 'diversified' scenario, new water is allocated to one of several pools as it enters the soil, and transfers between these pools are limited. Evaristo and colleagues' analysis¹ of soil-water data suggests that a diversified mode dominates worldwide: the distributions associated with plant water and evaporated water are shifted by similar amounts from the precipitation values, but the distribution associated with run-off is not shifted by much.

for — and raises many questions about — our understanding of water cycling. At a fundamental level, the diversified mode is conservative in that water is allocated to support multiple uses, with less water available than in the commingled model to support run-off or plant growth individually. The mechanisms that maintain the segregation of these pools of water as they move through the soil matrix remain unresolved, but understanding them is crucial for developing accurate models of soil-water partitioning. In particular, the relative roles of physical and temporal segregation remain unclear. Do plants draw water from different parts of the soil matrix from groundwater recharge, or do plant withdrawals happen at a different time from groundwater recharge?

The suggestion that plants that span many biomes have developed strategies of water use that focus on one pool of soil water largely to the exclusion of another is intriguing, given that drought stress is a major driver of plant mortality⁷. Perhaps the explanation lies in the transient nature of the soil-water stocks that contribute to recharge and run-off. Evaristo and co-workers' observation of the pervasive pattern of association between plant and soil water, but not with run-off-generating water, calls for further research. However, the isotopic methods used by the authors are less useful in the study of non-woody species such as grasses, so extending this work to some fast-growing and potentially less-discriminative water users will require new approaches.

The lack of water exchange between soil pools also calls into question some previous analyses of water-cycle processes (see ref. 8, for example), because it implies that methods for studying water partitioning that use measurements of chemical or isotope tracers in streams may be blind to the part of the soil-water balance sheet that involves plants and soil evaporation. Indeed, a study⁹ published earlier this year found that balancing the global water-isotope budget requires widespread hydrological separation in soils, consistent with Evaristo and colleagues' results, and called for a revision of previous global flux estimates from studies that did not consider hydrologic separation.

Finally, water allocation is inextricably coupled with soil biogeochemical reactions, from rock weathering to nutrient cycling, and the effect of diversified soil-water allocation on these processes may be enormous. Many reactions occur in thin films of water surrounding mineral grains. If these films are part of a long-lived soil-water pool, and there is little physical exchange of this water with through-flowing, run-off-generating water, what are the implications for the transfer of the reaction products to groundwater and streams? In this sense, the 'trickle-down' effects of soil-water economics may structure the entire soil

chemical system. Better understanding of the processes governing soil-water partitioning may ultimately help to resolve long-standing problems in geochemistry — such as discrepancies between field- and laboratory-based mineral-weathering rates¹⁰, which are central to our understanding of the global carbon cycle and to proposed geoengineering strategies for coping with anthropogenic emissions of carbon dioxide. ■

Gabriel Bowen is in the Department of Geology and Geophysics, University of Utah, Salt Lake City, Utah 84112, USA.
e-mail: gabe.bowen@utah.edu

MOLECULAR BIOLOGY

Unequal opportunity during class switching

The DNA breakage- and-repair mechanism that generates antibodies of different classes has, in theory, a 50% chance of occurring correctly. But this recombination turns out to be heavily biased towards productive events. SEE LETTER P134

JAVIER M. DI NOIA

Antibodies are proteins that recognize and neutralize invading pathogens. To accomplish this, antibodies (also called immunoglobulins) must access different tissues and recruit killer molecules and immune cells. These effector functions depend on the antibody's 'constant' region, which varies in protein sequence between the different classes of immunoglobulin — IgM, IgG, IgA and IgE. To produce antibodies appropriate to a particular infection, the class can be changed through an inducible genomic rearrangement called class-switch recombination (CSR)¹. This process can occur in two orientations, one of which results in a 'productive' rearrangement whereas the other prevents antibody production. Theoretically, these two events have an equal probability of occurring, which would give CSR a 50% failure rate that would limit the efficiency of antibody responses. But in this issue, Dong *et al.*² (page 134) demonstrate that CSR is highly non-random, with 90% of events resulting in a functional rearrangement.

Antibodies of the IgM class are the first to be produced when B cells of the immune system are stimulated by encounter with a pathogen. As the immune response progresses, the IgM constant region is replaced by another one, depending on the infection: IgG antibodies are effective against viruses and bacteria and are the antibodies induced by vaccination; IgA antibodies protect mucosal surfaces; and

1. Evaristo, J., Jasechko, S. & McDonnell, J. J. *Nature* **525**, 91–94 (2015).
2. Liang, X., Lettenmaier, D. P., Wood, E. F. & Burges, S. J. *J. Geophys. Res.* **99**, 14415–14428 (1994).
3. Neitsch, S. L., Arnold, J. G., Kiniry, J. R. & Williams, J. R. *Soil and Water Assessment Tool Theor. Document*. Version 2009 (Texas Water Resources Inst., 2011).
4. Beven, K. & Germann, P. *Water Resour. Res.* **18**, 1311–1325 (1982).
5. Brooks, J. R., Barnard, H. R., Coulombe, R. & McDonnell, J. J. *Nature Geosci.* **3**, 100–104 (2010).
6. Goldsmith, G. R. *et al. Ecohydrology* **5**, 779–790 (2012).
7. Anderegg, W. R. L. *et al. Proc. Natl Acad. Sci. USA* **109**, 233–237 (2012).
8. Jasechko, S. *et al. Nature* **496**, 347–350 (2013).
9. Good, S. P., Noone, D. & Bowen, G. *Science* **349**, 175–177 (2015).
10. Moore, J., Lichtner, P. C., White, A. F. & Brantley, S. L. *Geochim. Cosmochim. Acta* **93**, 235–261 (2012).

IgE antibodies attack certain parasites.

The constant regions defining the antibody classes are all encoded by the antibody heavy-chain gene (*Igh*), and the sequence for each is preceded by a distinct repetitive 'switch' (S) sequence — S μ , S γ , S α and S ϵ (Fig. 1). During CSR, the enzyme activation-induced deaminase (AID) causes simultaneous DNA double-strand breaks (DSBs) at the S μ and another S region¹. The DSBs, which can be up to 200 kilobases apart, are then joined by non-homologous end joining (NHEJ), a ubiquitous pathway for repairing broken chromosomes³. For productive CSR, the broken ends of the two separate DSBs must be joined in the orientation that places the new constant region in place of that of IgM and circularizes and deletes the intervening sequence (Fig. 1). Joining in the other orientation inverts the region between the DSBs, inactivating the antibody gene.

By adapting a high-throughput technique⁴ that enables analysis of the sequence and relative orientation of massive numbers of junctions between two DSBs, Dong *et al.* determined the relative frequency of deletion compared with inversion after inducing CSR. Their results show unambiguously that CSR is heavily biased towards the productive orientation, which they found in more than 90% of the joins between two switch regions broken by AID.

The authors further explored this finding by introducing into the genome of antibody-producing B cells sequences that are recognized

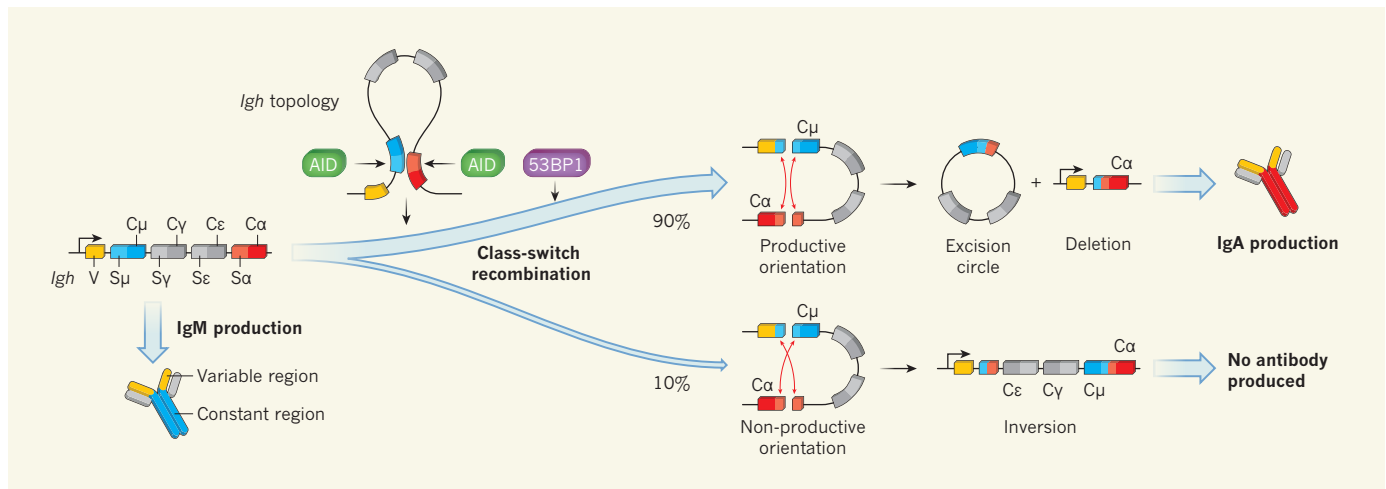


Figure 1 | Biased recombination. The *Igh* gene contains a region encoding the variable (V) region of an antibody (immunoglobulin) and several regions encoding the antibody's constant (C) regions, the latter each being preceded by switch (S) regions. The C region that is expressed determines the immunoglobulin class. The default class, IgM (encoded by C μ) can be changed to another class (IgG, IgE or IgA, encoded by C γ , C ϵ or C α) through class-switch recombination (CSR). In this process, the enzyme AID induces double-strand DNA breaks at the S μ region and at another S region. The broken

DNA is then rejoined through non-homologous end joining (red arrows). If the broken strands are oriented such that C μ is replaced by another constant region, the intervening DNA will circularize and be excised, and a productive antibody-encoding sequence is generated. If the strands are joined in an inverted orientation, no antibody is produced. Dong *et al.*² show that normal CSR is more than 90% biased towards the productive orientation. The authors present evidence suggesting that the topology of *Igh* during CSR and the DNA-repair factor 53BP1 are the primary contributors to this orientation bias.

by the enzyme I-SceI (a yeast enzyme that induces DSBs in DNA). They then looked at the orientation outcomes when one or both DSBs were made by I-SceI at a non-switch-region sequence, compared with breaks created by AID at an S region. It had been shown previously that replacing both S regions with I-SceI recognition sites allows CSR following expression of I-SceI⁵. Dong *et al.* find that I-SceI-initiated recombinations lack orientation bias, even when multiple I-SceI sites in tandem are used to mimic the repetitive nature of S regions. This lack of bias probably contributes to the relative inefficiency of CSR following I-SceI-induced DSBs compared with normal CSR⁵.

By contrast, the authors find that CSR between an I-SceI-induced non-S-region break and an AID-induced S-region break within the *Igh* gene is more than 75% biased towards the productive orientation. This result, together with the greater than 90% orientation bias of CSR between two AID-induced S-region breaks, suggests that the S regions and/or AID contribute to enforcing the productive orientation. However, although seemingly necessary, the S-region and AID are not sufficient for this bias — Dong *et al.* found no orientation preference when the AID-induced and the I-SceI-induced breaks were in two different chromosomes, a set-up that mimics AID-dependent chromosomal translocations⁶.

Previous work had shown that inter-chromosomal fusions between an I-SceI-induced and a spontaneous DSB are unbiased⁴. Future analysis of junctions between AID-induced or I-SceI-induced breaks in various genomic locations will make it possible to test

the authors' hypothesis that the spatial organization of the *Igh*, in which the S regions are brought into contact with each other within a physically restrained topological domain (Fig. 1), is a key determinant of the orientation preference of CSR. A recent paper shows that two S regions inserted into another gene targeted by AID (the immunoglobulin light-chain gene) are not joined to one another even if efficiently broken by AID⁷. Instead, the individual DSBs in each S region are just rejoined. This observation is consistent with a role for the specific topology of *Igh* in promoting the productive joining of breaks between S regions.

Dong *et al.* obtained further insight into the mechanism of CSR by studying cells lacking DNA-repair factors that are involved in the process. They found that orientation bias was reduced in the absence of the enzyme ATM kinase, which coordinates the response to AID-induced damage⁸, and in cells lacking the DNA-binding proteins H2AX, Rif-1 or 53BP1, which act to protect broken ends from being resected, thereby promoting NHEJ^{8,9}. The authors propose that inhibiting end resection accentuates an intrinsic predisposition of CSR to proceed in a specific orientation, dictated by the topology of *Igh*, by allowing NHEJ to repair breaks that are not correctly paired and could join in either orientation. However, 53BP1 prevents end resection by recruiting Rif-1 (ref. 9), and Dong *et al.* find that decreasing resection in 53BP1-deficient cells did not restore the orientation preference, thus revealing a resection-independent role for 53BP1 in determining orientation. Accordingly, 53BP1 is required for normal CSR but dispensable for I-SceI-mediated CSR^{3,8}. Putative 53BP1

functions include pairing of the S regions and influencing the topology of *Igh*⁹. Dissecting the role of 53BP1 in orientation bias will be one of the most interesting challenges arising from this work.

The only other example of orientation-biased DNA rearrangement is VDJ recombination, the process that assembles the antibody genes during B-cell development¹⁰. The mechanisms inducing this bias and that of CSR, as revealed by Dong and colleagues, are poorly understood and probably different. However, it is unlikely to be a coincidence that both have evolved to function in the most effective way to ensure the production of antibodies and thereby an efficient immune response. ■

Javier M. Di Noia is at the Institut de Recherches Cliniques de Montréal and Department of Medicine, Université de Montréal, Montreal, Quebec H2W 1R7, Canada.
e-mail: javier.di.noia@ircm.qc.ca

1. Stavnezer, J. & Schrader, C. E. *J. Immunol.* **193**, 5370–5378 (2014).
2. Dong, J. *et al. Nature* **525**, 134–139 (2015).
3. Alt, F. W., Zhang, Y., Meng, F.-L., Guo, C. & Schwer, B. *Cell* **152**, 417–429 (2013).
4. Chiarle, R. *et al. Cell* **147**, 107–119 (2011).
5. Zarrin, A. A. *et al. Science* **315**, 377–381 (2007).
6. Ramiro, A. R. *et al. Cell* **118**, 431–438 (2004).
7. Bonaud, A. E. L. *et al. Nature Commun.* **6**, 7613 (2015).
8. Daniel, J. A. & Nussenzweig, A. *Mol. Cell* **50**, 309–321 (2013).
9. Zimmermann, M. & de Lange, T. *Trends Cell Biol.* **24**, 108–117 (2014).
10. Helmink, B. A. & Sleckman, B. P. *Annu. Rev. Immunol.* **30**, 175–202 (2012).

This article was published online on 26 August 2015.

The quiet revolution of numerical weather prediction

Peter Bauer¹, Alan Thorpe¹ & Gilbert Brunet²

Advances in numerical weather prediction represent a quiet revolution because they have resulted from a steady accumulation of scientific knowledge and technological advances over many years that, with only a few exceptions, have not been associated with the aura of fundamental physics breakthroughs. Nonetheless, the impact of numerical weather prediction is among the greatest of any area of physical science. As a computational problem, global weather prediction is comparable to the simulation of the human brain and of the evolution of the early Universe, and it is performed every day at major operational centres across the world.

At the turn of the twentieth century, Abbe¹ and Bjerknes² proposed that the laws of physics could be used to forecast the weather; they recognized that predicting the state of the atmosphere could be treated as an initial value problem of mathematical physics, wherein future weather is determined by integrating the governing partial differential equations, starting from the observed current weather. This proposition, even with the most optimistic interpretation of Newtonian determinism, is all the more audacious given that, at that time, there were few routine observations of the state of the atmosphere, no computers, and little understanding of whether the weather possesses any significant degree of predictability. But today, more than 100 years later, this paradigm translates into solving daily a system of nonlinear differential equations at about half a billion points per time step between the initial time and weeks to months ahead, and accounting for dynamic, thermodynamic, radiative and chemical processes working on scales from hundreds of metres to thousands of kilometres and from seconds to weeks.

A touchstone of scientific knowledge and understanding is the ability to predict accurately the outcome of an experiment. In meteorology, this translates into the accuracy of the weather forecast. In addition, today's numerical weather predictions also enable the forecaster to assess quantitatively the degree of confidence users should have in any particular forecast. This is a story of profound and fundamental scientific success built upon the application of the classical laws of physics. Clearly the success has required technological acumen as well as scientific advances and vision.

Accurate forecasts save lives, support emergency management and mitigation of impacts and prevent economic losses from high-impact weather, and they create substantial financial revenue—for example, in energy, agriculture, transport and recreational sectors. Their substantial benefits far outweigh the costs of investing in the essential scientific research, super-computing facilities and satellite and other observational programmes that are needed to produce such forecasts³.

These scientific and technological developments have led to increasing weather forecast skill over the past 40 years. Importantly, this skill can be objectively and quantitatively assessed, as every day we compare the forecast with what actually occurs. For example, forecast skill in the range from 3 to 10 days ahead has been increasing by about one day per decade: today's 6-day forecast is as accurate as the 5-day forecast ten years ago, as shown in Fig. 1. Predictive skill in the Northern and Southern hemispheres is almost equal today, thanks to the effective

use of observational information from satellite data providing global coverage.

More visible to society, however, are extreme events. The unusual path and intensification of hurricane Sandy in October 2012 was predicted 8 days ahead, the 2010 Russian heat-wave and the 2013 US cold spell were forecast with 1–2 weeks lead time, and tropical sea surface temperature variability following the El Niño/Southern Oscillation phenomenon can be predicted 3–4 months ahead. Weather and climate prediction skill are intimately linked, because accurate climate prediction needs a good representation of weather phenomena and their statistics, as the underlying physical laws apply to all prediction time ranges.

This Review explains the fundamental scientific basis of numerical weather prediction (NWP) before highlighting three areas from which the largest benefit in predictive skill has been obtained in the past—physical process representation, ensemble forecasting and model initialization. These are also the areas that present the most challenging science questions in the next decade, but the vision of running

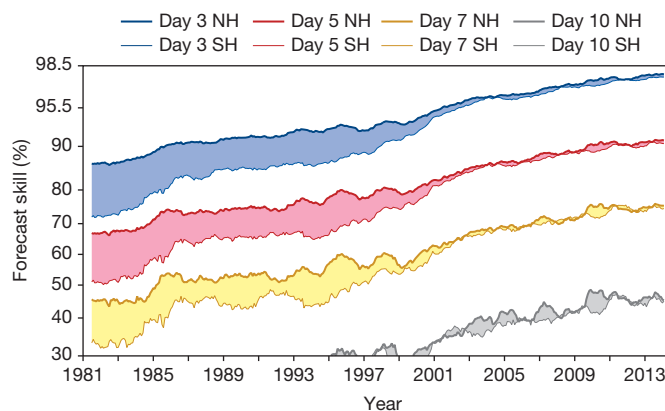


Figure 1 | A measure of forecast skill at three-, five-, seven- and ten-day ranges, computed over the extra-tropical northern and southern hemispheres. Forecast skill is the correlation between the forecasts and the verifying analysis of the height of the 500-hPa level, expressed as the anomaly with respect to the climatological height. Values greater than 60% indicate useful forecasts, while those greater than 80% represent a high degree of accuracy. The convergence of the curves for Northern Hemisphere (NH) and Southern Hemisphere (SH) after 1999 indicates the breakthrough in exploiting satellite data through the use of variational data¹⁰⁰.

¹European Centre for Medium-Range Weather Forecasts, Shinfield Park, Reading RG2 9AX, UK. ²Environment Canada, Trans-Canada Highway Dorval, Québec H9P 1J3, Canada.

global models at 1 km horizontal resolution, thus with an order of magnitude greater resolution than today, has added a new dimension, as it requires significant investment in high-performance computing with as-yet unknown technology.

The physics of forecasting

The Navier–Stokes and mass continuity equations (including the effect of the Earth’s rotation), together with the first law of thermodynamics and the ideal gas law, represent the full set of prognostic equations upon which the change in space and time of wind, pressure, density and temperature is described in the atmosphere⁴. These equations have to be solved numerically using spatial and temporal discretization because of the mathematical intractability of obtaining analytical solutions, and this approximation creates a distinction between so-called resolved and unresolved scales of motion. Physical processes that operate on unresolved scales down to the molecular enter the equations for the resolved scales through source terms for mass, momentum and heat originating from friction, moist processes such as condensation and evaporation, and radiative heating and cooling. Since these processes are typically unresolved they need to be ‘parameterized’ in terms of their interaction with the resolved scales. Simplifications can be applied that facilitate the numerical solution and reduce somewhat the complexity of the set of equations, as demonstrated for the first time—even though with limited success—by Richardson⁵. By introducing approximations that accurately describe the largest scales of motion in the atmosphere, the first attempt to use the first electronic computer for weather prediction was carried out in Princeton in 1950⁶. While the Princeton simulations were hindcasts, the first real-time forecasts were made in Stockholm in 1954⁷.

Only with increasing availability of supercomputing power in the 1970s was it feasible to solve the full set of equations as proposed by Abbe and Bjerknes⁸. Consequently, various numerical methods of solution emerged that addressed numerical stability, accuracy, computational speed⁹ and versatility to deal with more prognostic variables, and the interaction between resolved and unresolved scales¹⁰. The main components of these methods are: the representation of spatial variability by the choice of spatial discretization, the time stepping method, the treatment of boundaries, and the initialization approach¹¹. This capability has founded what we refer to as NWP¹². Today, a hierarchy of many models with different levels of complexity exists covering the full range between global climate projections¹³, global weather prediction, and local-area modelling for high-impact weather¹⁴ or air-quality prediction¹⁵.

Major steps

The improvements in the representation of unresolved processes in global models, the advent of ensemble methods producing forecast

uncertainty estimates, and the introduction of objective analysis techniques to determine the initial state have led to the predictive skill attained today. Representing physical processes, ensemble modelling and model initialization are also the key challenges for the future, combined with technological challenges associated with observations and computing, as we will discuss later.

Physical processes

Parameterizations capture radiative, convective and diffusive effects in the atmosphere and at the interface between the atmosphere and the surface, and are often determined by relatively small spatial scales^{16,17}. Figure 2 provides an illustration of these processes and where they are relevant. Despite not being resolved, these processes drive heat and momentum budgets at the grid scale^{18,19} and are crucial for achieving predictive skill. The degree of parameterization and therefore the representation of the basic physics varies significantly for different processes²⁰. For example, the global model formulation for radiation and cloud microphysics processes is similar to that used in regional and high-resolution models because the formulation accounts for the basic small-scale physics, which is similar across these model spatial scales, even if they require added complexity going to higher spatial resolution. The formulations are mostly limited by our understanding of physical process detail needed for parametric representations that define the spatially averaged impact of the process on momentum and heat fluxes. On the other hand, deep convection and specific boundary layer processes require a higher degree of parametric formulation as they only occur in small fractions of the grid scale; consequently these parameterizations critically depend on which resolution is actually used.

Parameterizations play a fundamental role in determining predictive skill because they determine key aspects of the simulated weather, such as clouds and precipitation, as well as temperature and wind. In operational NWP models, essentially the same formulation for the parameterizations is used for scales of 10–100 km in short-to-medium range forecasts, minimization algorithms used for model initialization, and seasonal range forecasts. Achieving this element of ‘grid-scale invariance’ while including as much physical process detail as possible has been a fundamental breakthrough in the recent past.

Ensemble modelling

Early in the twentieth century, Poincaré²¹ recognized that forecasts of nonlinear systems can vastly differ if small perturbations are applied to the initial conditions, and that this difficulty could be fundamental in limiting predictive skill. In the 1950s, Thompson performed one of the first quantitative estimates of initial errors growing during the forecast²², while Lorenz²³ formulated this understanding more holistically and

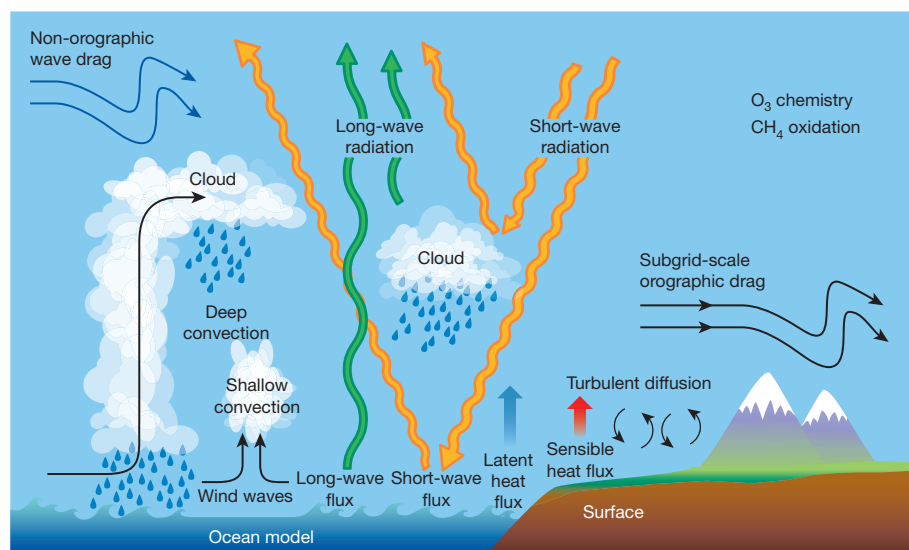


Figure 2 | Physical processes of importance to weather prediction. These are not explicitly resolved in current NWP models but they are represented via parameterizations describing their contributions to the resolved scales in terms of mass, momentum and heat transfers.

founded chaos theory as a result of his attempt to quantify atmospheric predictability. From his conclusion—that unstable systems have finite, state dependent limits of predictability—was born the need for encapsulating the growth of initial condition uncertainties, their evolution as a function of the atmospheric state, and errors introduced by imperfect models. The recognition of imperfect forecasts²⁴ and determining how to calculate analysis and forecast uncertainty using an ensemble approach²⁵ represent major and unique accomplishments in physical sciences. This is particularly true for the prediction of highly variable parameters like precipitation (Fig. 3), where ensemble spread quantifies forecast uncertainty of rainfall location and intensity and thus provides essential information to users.

The nonlinear complexity of the system means that purely statistical methods to assign an uncertainty to the forecast are inadequate. Instead, an ensemble of many complete, physical, nonlinear realizations of the system is needed^{26,27}, providing a seamless analysis and forecast ensemble in which observational information is used to reduce uncertainty. In practice, the ensemble members are created using perturbations, equivalent to analysis and model errors, added to the initial state and the model physical processes. Determining these perturbations consistently and seamlessly so that the ensemble provides a good estimate of uncertainty across a wide range of prediction scales is challenging, and the input of mathematics and statistical physics expertise was crucially important^{28,29}. Weather forecasts today involve an ensemble of numerical weather predictions, providing an inherently probabilistic assessment.

Model initialization

Early methods for the specification of initial conditions were based on the analysis of graphical and synoptic weather charts. Various forms of interpolation procedures were later replaced by data assimilation techniques based on optimum control theory³⁰. The derivation of the current state (called the analysis) of the atmosphere and surface is treated as a Bayesian inversion problem using observations, prior information from short-range forecasts and their uncertainties as constraints as well as the forecast model^{31,32}. These calculations, involving a global minimization, are performed in four dimensions to produce an analysis that is physically consistent in space and time and can deal with huge amounts of observational data that are heterogeneously distributed in space and time (such as the vast amount and diversity of satellite data used for Earth observation since the 1980s). Since initial state uncertainty estimation is also crucial for ensemble prediction and because data assimilation employs both imperfect observations and forecast model, ensemble methods have also become an integral part of data assimilation³³, as shown in Fig. 4.

The operational implementation of these four-dimensional variational (4D-Var) data assimilation techniques³⁴ marks a major milestone in operational global NWP. At the European Centre for Medium-Range Weather Forecasts (ECMWF) this occurred in 1997³⁵, followed by

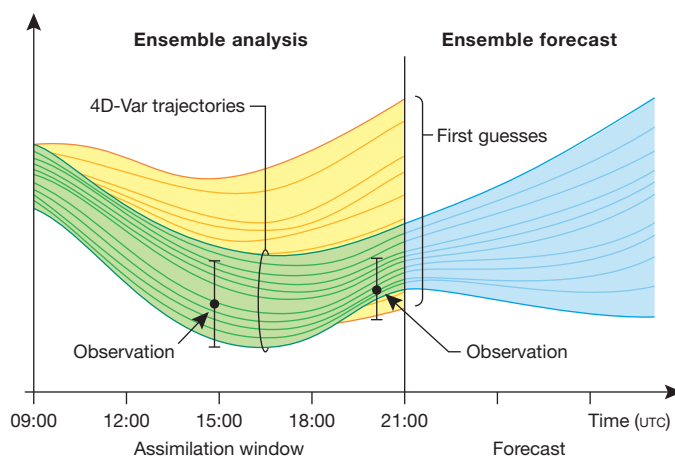


Figure 4 | Schematic of the ensemble analysis and forecast cycle. Global ensemble forecast trajectories, which have been initialized by a previous analysis ensemble, are produced over a time window (for example, 09:00–21:00 UTC). These provide estimates of the current weather (first guesses). The difference between these forecasts and available observations (shown as data points with error bars) is the short-range forecast error. By minimization in four dimensions employing variational techniques, improved estimates (4D-Var trajectories) are created with reduced distance to observations. The next cycle of ensemble forecasts is then initialized from these refined analyses. Image courtesy of M. Bonavita (ECMWF).

Météo-France in 2000³⁶, the Met Office in 2004³⁷, both the Japan Meteorological Agency³⁸ and Environment Canada in 2005³⁹, and the United States Naval Research Laboratory in 2009⁴⁰. Development and first implementation of 4D-Var took more than 10 years, and further research has substantially refined the main ingredients. These were the increasing use of satellite radiance data by combining the forecast model with computationally efficient radiative transfer models^{41,42}, the much refined characterization of short-range forecast⁴³ and observation errors⁴⁴ using state dependent weights for each, and better use of observations arising from significant improvements of physical parameterizations⁴⁵.

Predictability and predictive skill

A continuing and important area of research focuses on the sources of predictability in the Earth system. Forecasting future weather is like a battleground, with the forces of predictability pitched against those of unpredictability. The sources of predictability include large-scale forcing of smaller-scale weather, teleconnections or the chain of predictability across different geographical areas⁴⁶, and the interactions between atmosphere, land surfaces and vegetation, sea-ice and ocean acting on longer timescales. The sources of unpredictability include

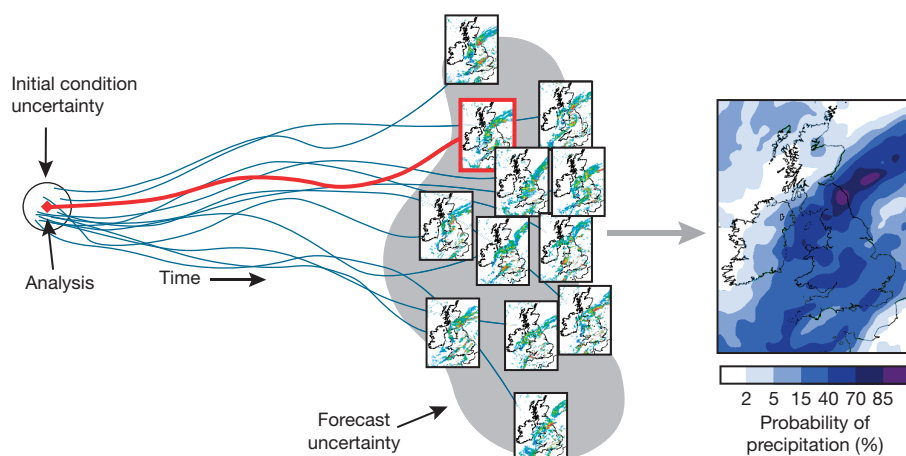


Figure 3 | Schematic diagram of 36-h ensemble forecasts used to estimate the probability of precipitation over the UK. A single forecast (red frame, centre) is generated by integrating the model forward in time from the analysis of initial atmospheric state (left). Small perturbations to the analysis, within known analysis uncertainty, provide an ensemble of forecast solutions, which sample the forecast uncertainty (multiple frames). These solutions are combined, including some spatial neighbourhood sampling, to provide a smooth estimate of probability of precipitation (right). Image courtesy of K. Mylne (Met Office).

instabilities injecting chaotic ‘noise’ at small scales and the upscale propagation of their energy, the errors associated with numerical and physical approximations, as well as the insufficient number and poor use of observations. Box 1 provides an example of such teleconnections and the sources of poor forecast performance over Europe in the medium range.

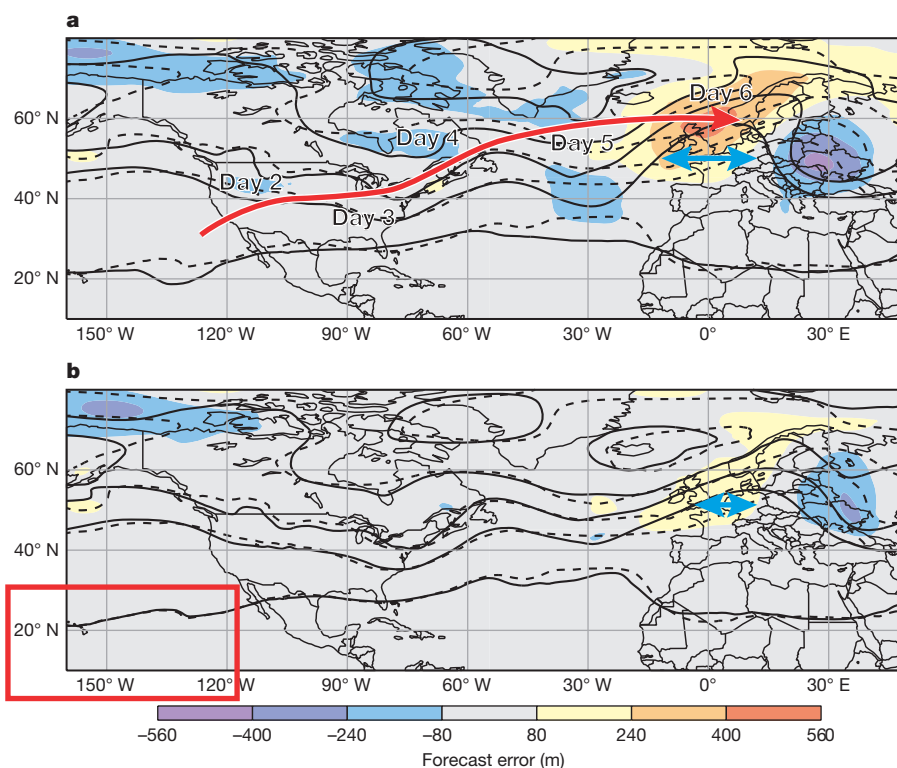
The outcome of this ‘battle’ can be described as noise growing nonlinearly during the forecast and thereby leading to fundamental limits of how far into the future certain structures can be predicted. The limit for small-scale events is between hours and days, for accurate and reliable prediction of high-impact weather events about 1–2 weeks, for prediction of large-scale weather patterns and regime transitions about a month, and for global circulation anomalies about a season⁴⁷. The longer the forecast range the more the predictive skill relates only to anomalies, that is, the difference between the state and its modelled climatological mean, and the more important space-time averaging becomes to identification of the signal. In the short range predictive skill exists for the details, while in the long range skill relates to larger-scale structures. Predictive capability that is seamless across this wide range of forecast

horizons is therefore about capturing processes acting on very different time and space scales.

NWP has a fundamental advantage over many other scientific disciplines in that its skill is objectively evaluated daily and globally, so that success and failure of forecasts is accurately known and pathways to improve predictive skill can be effectively tested^{48,49}. To evaluate forecast skill, metrics such as mean and root-mean-square errors, and the correlation of the forecast with analysis anomalies of upper-air and surface forecast fields are used. In addition, scores targeting more variable parameters such as precipitation⁵⁰ exist. Model biases become significant further into the forecast range. While biases can be reduced through calibration using past forecasts⁵¹, the identification of their sources in complex models remains one of the dominating challenges for NWP and even more so for climate prediction⁵². Diagnostic methodologies employing data assimilation statistics⁵³ can help since the signature of most biases is already evident in the analysis and early in the forecast, even though their magnitude is small. This approach offers benefits for weather and climate science alike.

BOX 1

Sensitivity of forecasts to initial conditions and error propagation



Box 1 Figure | Maps showing the long-range impact of model initialization on the European forecast. Panel **a** shows the day-6 mean forecast error (the height of the 500 hPa pressure level in metres) of the flow at around 5 km height (colour-coded shading), the forecast itself (solid isolines) and the verifying analysis (dashed isolines) valid on 15 February 2014. Over the western US, the jet stream extended far to the south, aligned with a lower level trough. The long red arrow indicates the travel path of an atmospheric wave disturbance guided by the westerly flow. The presence of a large-scale dipole error pattern highlights the lag between forecast and analysed state (blue double-headed arrow). The large forecast errors over Europe were mostly produced by a phase-shift of the wave that increased with time. Back-tracking the wave propagation path

identifies the tropical East Pacific (boxed in **b**) as a likely location of a possible forecast error source. This area was characterized by very large 24-h forecast errors of upper-level winds because of the paucity of wind observations there. When running an experiment where the area in the box in **b** is relaxed towards the analysis rather than evolving in the forecast, the strong initial growth of forecast errors is reduced and, six days later, the lag of the wave patterns between forecast and analysis is reduced over Europe (blue double-headed arrow), producing about half of the original forecast errors. This experiment demonstrates the long-range impact of model initialization, the linkage between tropics and mid-latitudes, and thus shows an example of how predictive skill in the one-week time range can be increased.

As NWP involves an ensemble of forecasts, evaluation metrics need to assess the moments of probability distributions such as ensemble mean error and the sharpness of the distributions. Forecast reliability is determined by comparing forecast distributions with the observed frequency of occurrence. Since ensembles are designed to provide valuable information on the probability of weather extremes⁵⁴, scores targeting the tails of probability distributions are being developed accounting for sparse statistics⁵⁵.

In addition, comprehensive feature-based evaluation is available for tropical cyclones⁵⁶ or weather regimes⁵⁷, and for the evaluation of how well models represent the links between lower and higher latitudes^{58,59}, troposphere and stratosphere^{60,61}, planetary wave activity driving synoptic scale features⁶², and synoptic scales interacting with small-scale convection^{63,64} and the surface^{65,66}.

An effective way to verify predictive skill also arises from combining weather with hydrological modelling, whereby predicted river streamflow and discharge help to evaluate predictions of precipitation, run-off and storage in NWP models, both for single realization and ensemble forecasts^{67,68}. The enhancement of weather models with variables describing atmospheric composition such as aerosols and trace gases also introduces new ways to evaluate atmospheric evolution by considering tracer advection and model chemistry parameterizations⁶⁹.

Where we are today

Operational NWP centres provide predictions from the very short range at kilometre scale multiple times per day up to global seasonal forecasts at tens of kilometres horizontal resolution once per month. These forecasts relate to the weather but are also extending to air-quality⁷⁰ and hydrological⁷¹ applications.

Data assimilation algorithms employ the forecast model and of the order of 10^7 observations per day to derive initial conditions that are physically consistent in four dimensions: over the globe, from surface up to mesosphere (~ 80 km) and along time windows from hours to days. Operational models are updated frequently to incorporate new science that enables improvements in the representation of model physics and model uncertainty, in numerical algorithms and observational data usage, and to enhance computational efficiency.

Gauging the relative contributions to success and progress from model development, data assimilation algorithms and observational data usage is difficult because they are interdependent. More accurate model physics means that forecasts compare better with observations and facilitate improved data assimilation; in turn this permits ingestion of more observations thereby further improving forecasts.

NWP has also benefited enormously from computing advances. In terms of floating point operations, computing power has increased by about one order of magnitude every five years since the 1980s. This is the result of processor technology advances and more processors being used. Intel co-founder Gordon Moore's law states that computing power doubles every 18 months owing to increased transistor density per chip and clock speeds. This growth has gone hand-in-hand with the increasing size of the analysis and forecast computational task in NWP. At ECMWF, the data assimilation performs model integrations in multiple stages totalling of the order of 100 iterations across a 12-h window for a total of 650 million grid-point calculations. In parallel, about 10 million radiance calculations are performed to compare the forecast model with satellite observations from more than 60 instruments. Today, the ECMWF 16-km highest-resolution model performs calculations on two million grid columns with 10-min time stepping over a 10-day period, that is, 1,440 time steps. The corresponding ensemble produces 15–30 day forecasts with 50 members with a horizontal resolution of 30–60 km and 30-min time steps. Thus twice per day about 40 billion grid-column calculations are performed in about 2.5 h real time. This computing task demands some of the largest supercomputing facilities available.

The time series describing the improving skill of global NWPs is impressive (Fig. 1), revealing that while there is some year-to-year variability, for more than three decades forecast skill has been advancing

continuously^{72,73}. Predictive skill improves at a rate such that useful skill is retained one more day into the forecast range for every decade of research and development. This steady progress has been the result of advances in the science, in the utilization of observations and in supercomputing capacity. Some of the fluctuations in skill are a result of periods when the atmosphere exhibits more or less potential predictability. This means that certain weather regimes appear to be easier to predict accurately further into the future than others. Our understanding of these regimes of flow is developing and enabling a more discerning quantification of predictive skill to be developed.

The future is bright

The evolution of weather science as well as of high-performance computing and observing systems in the future is crucial for continuing the progress in NWP. Critical scientific and technological cross-roads have been reached or are very likely to be reached in the near future. Consequently, the present period is of fundamental importance for how weather forecasting and also climate science will evolve. Building on anticipated advances in the understanding of physical processes, in numerical model development, in observation technology and high-performance computing, the vision for global weather and climate modelling a decade or more in the future is as follows: in terms of resolution to be able to perform global convection-resolving simulations at a horizontal resolution of the order of 1 km; in terms of complexity to be able to run fully coupled atmosphere–land–ocean–sea-ice models. Ensembles at this resolution and complexity will predict probabilities of dynamics, physics, chemistry and probably selected bio-chemical processes into the multi-seasonal range for weather, and into the multi-decadal range for climate. These global predictions provide essential initial and boundary information for finer-scale limited-geographical-domain simulations of short-range detailed weather development.

The scientific challenges

The main scientific challenges for future global NWP relate to the main themes that have produced key advances in the recent past and that have brought weather forecasting to the level where it is today: physical process parameterization, analysis and forecast uncertainty formulation through ensembles, and the provision of physically consistent initial conditions for forecasts using observations. There are a number of key areas in which substantial progress can be expected in the future that also require significant advances compared to current thinking.

Regarding physical parameterizations, one might anticipate that with increasing resolution the need for parameterization would be gradually reduced. For radiation and cloud processes and land surface models this is a matter of moving current schemes towards fully explicit models already used in regional and local applications at the kilometre scale. For convection, the situation is more complex because large tropical convective clouds or organized convection occur even at currently resolved scales (15 km) while embedded small-scale convective plumes may not be resolved even at 1 km and will still require parameterization. This range of model resolutions with partly resolved convection is also referred to as the grey zone, since resolved and parameterized contributions to fluxes need to be quantified and combined. Existing schemes assume that convection is entirely unresolved and so they are not able to adequately represent the impact of both resolved and unresolved process components on heat and momentum at resolved scales in the grey zone.

High-resolution limited-area cloud models have demonstrated that the dynamic modes of organized convection can be captured and that the modelling of the lifecycle of convection, cloud organization or its interaction with large-scale circulation can be improved⁷⁴. Whether running global models at scales of the order of 1 km also eliminates all convection-related uncertainties and produces a fundamental stepping stone for reduced model biases and enhanced predictive skill at all forecast ranges is not clear at present⁷⁵. As these high resolutions are not yet in reach, convection parameterizations will remain crucial for global

weather and climate modelling for the next decade⁷⁶ and progress in this area will require joint efforts in the weather and climate communities^{77,78}.

There are two other areas that need much more attention in the future and promise significant boosts of skill, but also involve substantial investments in scientific development and computing.

First, the uncertainties inherent to physical parameterizations, either from incomplete process understanding or the dilemma of representing the impact of unresolved processes on the resolved scales, may require a fundamentally different approach. Elements of parameterizations or entire schemes are likely to require components that appear statistical to the large scales because they are not fully determined by the resolved scales⁷⁹. Examples are stochastic sampling of parameter probability distribution functions, stochastically driven sub-cell models, or super-parameterizations⁸⁰ through embedding entire convection-resolving simulations at sub-grid scale. How radical this approach needs to be is currently not clear.

Second, more physical as well as chemical processes will be added. More physical processes will be needed because of the modelled coupling of the atmosphere with ocean, land surface and sea-ice models, some of which are already in operational use today^{81,82}. Each coupling has its own characteristic space and time scales and the coupling *per se* provides most benefit beyond the 3–7 day range since ocean, sea-ice and land surface processes are relatively slow and mostly affect longer-term system memory. However, there are examples where coupling also affects the short range: for example, when oceanic upwelling in the wake of slowly moving tropical cyclones affects their intensity, or where rainfall over land is strongly constrained by surface evaporation and thus soil moisture.

The greatest scientific challenge for coupling is associated with matching fluxes at the interfaces where systematic errors in each component interact⁸³ and can produce model shocks and compensating changes of mean state at every coupling time step and through feedbacks in longer integrations.

Atmospheric constituents such as trace gases and aerosols directly affect radiative heating, but aerosols can also act as condensation nuclei in cloud formation and heterogeneous chemistry occurs at the surface of polar stratospheric clouds, accelerating ozone destruction. Nevertheless, aerosols and trace gases are important to forecast in their own right because of their impacts on air quality. An associated challenge from adding more physical and chemical processes is that initial conditions for these constituents are also required and thus more and complex observations need to be assimilated. Ensemble prediction reliability beyond the medium range will therefore be enhanced by representing the uncertainty of much more complex processes in models and by being able to initialize coupled models using much more diverse observations.

Using more of the existing and new observations, and advances in data assimilation pose more science challenges for NWP. Currently, each global forecast uses about 5–10% of the total satellite data volume; this fraction contains most of the information content for that particular forecast. This approach is of fundamental importance to optimally manage the substantial global investment in Earth observation, especially from satellites⁸⁴. However, NWP is limited by insufficient observational data. Beyond the maintenance of the backbone satellite and ground-based observing systems that measure vertical profiles of temperature, moisture, clouds and near-surface weather, fundamental observables are missing. An example is the direct observation of upper-level wind with Doppler-radar technology⁸⁵, but this technology is not yet available in operational satellite programmes. Wind information is primarily needed in the tropics, an area covering around 50% of the Earth and where sparse observations are a serious impediment to increased analysis accuracy. However, the existing backbone observations also need to be provided by a robust and resilient observing system, which requires substantial international investment and coordination. A similar level of coordination is required for satellite and ground-based observations.

Notwithstanding the complexity of current data assimilation there are many challenges for the future, most of all regarding improved solution

algorithms; such algorithms will be targeted at enhancing the exploitation of new observational data, but will also be able to handle improved models. Computational affordability will continue to be a constraint, given that a sizeable proportion of the cost of producing a forecast is associated with data assimilation. Next-generation data assimilation methods will probably employ fundamentally new mathematics, but assimilation methods in the near future will probably be based on a combination of existing concepts.

Current algorithms commonly rely on linear principles and variational methods, whereas certain components, such as error statistics, are obtained from ensembles. The variational principle has been implemented in different ‘flavours’ and the next decade is likely to be dominated by either choosing the most effective combination of variational and ensemble elements⁸⁶ or by using purely ensemble based methods like ensemble Kalman filters⁸⁷. Smaller-scale effects operating on shorter time scales (for example, convection) may require nonlinear data assimilation methods for which only limited experimentation with idealized models exists⁸⁸. These are currently difficult to generalize for global operational applications.

Coupled data assimilation will become critical for the initialization of the future coupled models⁸⁹. This assimilation will need to include atmospheric composition (aerosols, trace gases) as well as ocean, land surfaces and sea-ice. Each Earth-system component has particular process characteristics and space–time scales, and dealing with those in a fully unified data assimilation framework will be extremely challenging.

Technological challenges

Today’s highest-performance computers employed in NWP rank in the top 20 of the 500 most powerful systems and execute computations at petaflop (10^{15} floating point operations) per second rates, ingesting of the order of 100 Mbytes of observational data and producing of the order of 10 Tbytes (that is, 10×10^{12} bytes) of model output per day. Future generations of global NWP models with kilometre scales in the horizontal will integrate of the order of 100 prognostic variables over about 5×10^8 grid points for of the order of 100 ensemble members with time steps of seconds in an atmosphere with about 100 levels, coupled to surface models of somewhat smaller dimensions. Observational data usage will also increase by an order of magnitude owing to the internationally coordinated availability of high-resolution spectrometers in low-Earth and geostationary orbits with thousands of spectral channels.

However, the expected future high-performance computing technology development will impose new constraints on how to address the science challenges. In the past, processor performance has evolved according to Moore’s law⁹⁰, as has memory capacity and processor clock-speeds. This trend cannot be expected to continue in the future as energy cost has to be reduced. In the future, much more emphasis will be placed on parallel computing and this is where the ‘scalability’ of an application becomes important, providing time-to-solution gains when the model is run on more (and combinations of different types of) processors. The gain from the parallel execution of parts of the code is limited by the sequentially run elements, which fundamentally limits scalability, as does the need to exchange large amounts of data between processors. Making NWP codes more scalable is among the top priorities in NWP for the next 10 years.

For NWP centres such as ECMWF, the upper limit for affordable power usage may be about 20 MVA (ref. 91). The likely future NWP system will be of the order of 100–1,000 times larger as a computational task than today’s systems, and would require about 10 times more power. Figure 5 illustrates the increase in compute cores and electric power supply if model resolution is increased for a single forecast and a 50-member ensemble, assuming today’s model design and available technology. To approach the resolutions of 1–5 km that are considered crucial for resolving convection, high-performance computers of unprecedented dimension and cost (assuming the use of conventional technology) would be required.

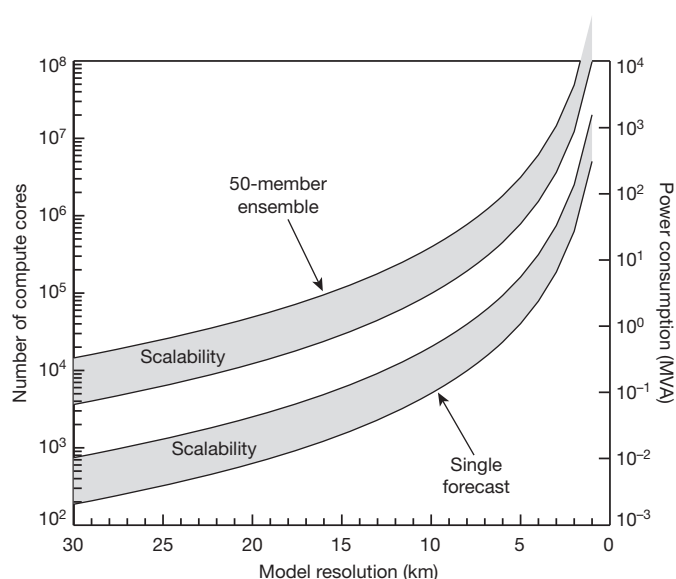


Figure 5 | CPU and power requirements as a function of NWP model resolution. Simplified illustration of the number of compute cores (left y-axis) and power (in units of megawatt, MVA, right y-axis) required for single 10-day model forecast (lower curves) and 50-member ensemble forecast (upper curves) as a function of model resolution, given today's model code and compute technology. The shaded area indicates the range covered when assuming perfect scaling (bottom curve) and inefficient scaling (top curve), respectively. Today's single global forecasts operate at around 15 km while ensembles have around 30 km resolution.

A change of paradigm is therefore needed regarding hardware, design of codes, and numerical methods⁹². New technologies will combine and integrate low-power processors with the successors of today's CPUs to give the best of both worlds—namely, highly parallel compute performance with little data communication at lower clock rates, and CPU-type performance with large memory, a fast data interface and higher clock rates. Code design and algorithm choice must be adapted to this technology by optimizing floating point operation counts and memory usage, which is a fundamental challenge given that we are dealing with vast heritage codes with millions of lines of instructions. In 10 years, global ensemble forecasts will be run on the order of 10^5 – 10^6 processors. Fault awareness and resilience management will be crucial, given the certainty of processor failures and the advent of inexact low-energy hardware⁹³.

This computing challenge is enhanced by the requirements for data distribution and archiving. While data growth appears slower than compute growth, exabyte (10^{18} bytes) data production may be reached earlier than exaflop computing. Re-computing is even more costly than archiving, and thus it is inevitable that the data challenge will need to be tackled with high priority⁹⁴. As for future processor technology, hardware will limit data transfer bandwidth. Occasional hardware failure needs to be actively accounted for by designing resilient storage systems. Such failures also have fundamental implications for the design of future work flows. Advanced data compression methods need to be implemented, and standardized and supported by the weather and climate community.

Many technological opportunities and challenges will arise from future Earth observing systems. At the high end, new satellite instrument technology will increasingly move towards hyper-spectral radiometers, with thousands of spectral channels sounding the atmospheric thermodynamical state and composition, together with active instruments (such as high-resolution radars and lasers) sounding surface characteristics, aerosols, wind, water vapour, clouds and precipitation. Both instrument categories can produce data rates of the order of 100 Gbytes per day that require downlinks, pre-processing, data dissemination within a few hours and ingestion in forecasting systems. The distribution and archiving of

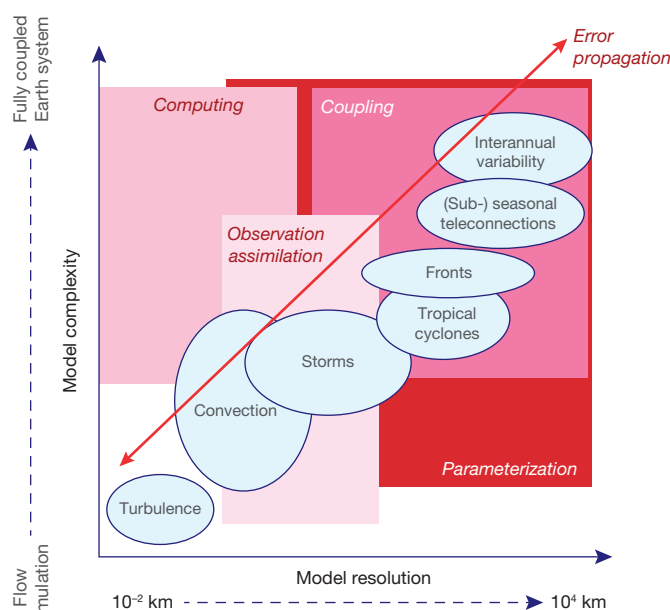


Figure 6 | Key challenge areas for NWP in the future. Advances in forecast skill will come from scientific and technological innovation in computing, the representation of physical processes in parameterizations, coupling of Earth-system components, the use of observations with advanced data assimilation algorithms, and the consistent description of uncertainties through ensemble methods and how they interact across scales. The ellipses show key phenomena relevant for NWP as a function of scales between 10^{-2} and 10^4 km resolved in numerical models and the modelled complexity of processes characterizing the small-scale flow up to the fully coupled Earth system. The boxes represent scale-complexity regions where the most significant challenges for future predictive skill improvement exist. The arrow highlights the importance of error propagation across resolution range and Earth-system components.

these data volumes will need to be managed with a similar parallelized approach as the model output. Data dissemination will only be feasible if compression techniques are applied, potentially accepting 'information loss'⁹⁵. At the low end, the use of commodity devices, such as mobile phones, with good sampling but less accuracy for gathering meteorological observations is only starting now, but offers potential for high-density observational networks in certain areas^{96,97}.

It is clear that scientific and technological challenges are interdependent in many areas. The efficiency of computing and data handling imposes hard limits on model complexity in weather and climate models that are run within tight production schedules, and it will be challenging to run globally at 1 km convection-resolving scales. This trade-off between scientific and compute performance is not new, but 'scalability' issues add a new dimension⁹⁸.

The quiet revolution of numerical weather prediction has required combined scientific, observing and computational technology advances to be made. This combination is common to all natural sciences that necessitate the solution to large problems, such as simulating the neurological connectivity of the human brain or the evolution of the galaxies in the cosmos. Further advances require more interdisciplinary research at the science–technology interface. As society's requirement for more accurate and reliable information regarding weather and climate grows ever more pressing, global numerical models will need to increase in both resolution and complexity. This further progress in global NWP can be made but will require combined investment in all the elements reviewed in this paper⁹⁹, as summarized schematically in Fig. 6.

Received 19 February; accepted 16 July 2015.

1. Abbe, C. The physical basis of long-range weather forecasts. *Mon. Weath. Rev.* **29**, 551–561 (1901).

2. Bjerknes, V. Das Problem der Wettervorhersage betrachtet vom Standpunkt der Mechanik und Physik. *Meteorol. Z.* **21**, 1–7 (1904).
3. Lazo, J. K., Morss, R. E. & Demuth, J. L. 300 billion served: sources, perceptions, uses, and values of weather forecasts. *Bull. Am. Meteorol. Soc.* **90**, 785–798 (2009).
4. Kalnay, E. *Atmospheric Modeling, Data Assimilation and Predictability* (Cambridge Univ. Press, 2003).
- A classic textbook on the fundamentals of modern weather prediction.**
5. Richardson, L. F. *Weather Prediction by Numerical Process* (Cambridge Univ. Press, 1922).
- Already in the early 1920s the author proposed numerical algorithms to solve the NWP partial differential equations using a huge staff of humans as computers.**
6. Charney, J. G., Fjortoft, R. & Neumann, J. v. Numerical integration of the barotropic vorticity equation. *Tellus* **2**, 237–254 (1950).
- The first calculation of a numerical weather prediction on an electronic computer.**
7. Bolin, B. Numerical forecasting with the barotropic model. *Tellus* **7**, 27–49 (1955).
8. Lynch, P. The origins of computer weather prediction and climate modeling. *J. Comput. Phys.* **227**, 3431–3444 (2008).
9. Robert, A. J. A semi-Lagrangian and semi-implicit numerical integration scheme for the primitive meteorological equations. *J. Meteorol. Soc. Jpn* **60**, 319–324 (1982).
- This seminal paper presents a numerical method used worldwide since the 1990s to solve the partial differential equations with a significantly longer time step thus greatly enhancing efficiency.**
10. Staniforth, A., Wood, N. & Côté, J. A simple comparison of four physics–dynamics coupling schemes. *Mon. Weath. Rev.* **130**, 3129–3135 (2002).
11. Williamson, D. L. The evolution of dynamical cores for global atmospheric models. *J. Meteorol. Soc. Jpn* **85**, 241–269 (2007).
12. Lorenz, E. N. Reflections on the conception, birth and childhood of numerical weather prediction. *Annu. Rev. Earth Planet. Sci.* **34**, 37–45 (2006).
13. Flato, G. *et al.* in *Climate Change 2013* (eds Stocker, T. F. *et al.*) 741–866 (Cambridge Univ. Press, 2013).
14. Dudhia, J. A history of mesoscale model development. *Asia-Pac. J. Atmos. Sci.* **50**, 121–131 (2014).
- Reviews the evolution of limited area modelling.**
15. Zhang, Y. Online-coupled meteorology and chemistry models: history, current status, and outlook. *Atmos. Chem. Phys.* **8**, 2895–2932 (2008).
16. Arakawa, A. Adjustment mechanisms in atmospheric motions. *J. Meteorol. Soc. Jpn* **75**, 155–179 (1997).
17. Williams, P. D. Modelling climate change: the role of unresolved processes. *Phil. Trans. R. Soc. A* **363**, 2931–2946 (2005).
18. Grenier, H. & Bretherton, C. S. A moist PBL parameterization for large-scale models and its application to subtropical cloud-topped marine boundary layers. *Mon. Weath. Rev.* **129**, 357–377 (2001).
19. Arakawa, A. The cumulus parameterization problem: past, present, and future. *J. Clim.* **17**, 2493–2525 (2004).
- Reviews the necessity and challenges of parameterizing sub-grid scale physical processes for cumulus clouds.**
20. Stensrud, D. J. *Parameterization Schemes, Keys to Understanding Numerical Weather Prediction Models* (Cambridge Univ. Press, 2007).
21. Poincaré, H. *Science and Method* (T. Nelson, London, 1914).
22. Thompson, P. D. Uncertainty of initial state as a factor in the predictability of large scale atmospheric flow patterns. *Tellus* **9**, 275–295 (1957).
23. Lorenz, E. N. Deterministic non-periodical flow. *J. Atmos. Sci.* **20**, 130–141 (1963).
- One of the most influential papers establishing the fundamentals of chaos theory applied to numerical weather prediction.**
24. Slingo, J. & Palmer, T. N. Uncertainty in weather and climate prediction. *Phil. Trans. R. Soc. A* **369**, 4751–4767 (2011).
25. Zhang, H. & Pu, Z. Beating the uncertainties: ensemble forecasting and ensemble based data assimilation. *Adv. Meteorol.* **2010**, 432160 (2010).
26. Epstein, E. S. Stochastic-dynamic prediction. *Tellus* **21**, 739–759 (1969).
27. Leith, C. E. Theoretical skill of Monte Carlo forecasts. *Mon. Weath. Rev.* **102**, 409–418 (1974).
- Discusses the estimation of atmospheric forecast uncertainties using a Monte Carlo approach based on an ensemble of perturbed numerical predictions.**
28. Ehrendorfer, M. Predicting the uncertainty of numerical weather forecasts: a review. *Meteorol. Z.* **6**, 147–183 (1997).
29. Palmer, T. N. Towards the probabilistic Earth-system simulator: a vision for the future of climate and weather prediction. *Q. J. R. Meteorol. Soc.* **138**, 841–861 (2012).
30. Lions, J. *Contrôle Optimal de Systèmes Gouvernés par des Équations aux Dérivées Partielles* (Dunod-Gauthier-Villars, Paris, 1968).
31. Lorenc, A. C. Analysis methods for numerical weather prediction. *Q. J. R. Meteorol. Soc.* **112**, 1177–1194 (1986).
32. Daley, R. *Atmospheric Data Analysis* (Cambridge Univ. Press, 1991).
- A comprehensive textbook on data assimilation and its application in atmospheric sciences.**
33. Navon, I. M. in *Data Assimilation for Atmospheric, Oceanic and Hydrologic Application* (eds Park, S. K. & Xu, L.) 21–65 (Springer, 2009).
34. Courtillot, P., Thépaut, J.-N. & Hollingsworth, A. A strategy for operational implementation of 4D-Var, using an incremental approach. *Q. J. R. Meteorol. Soc.* **120**, 1367–1387 (1994).
35. Rabier, F., Järvinen, H., Klinker, E., Mahfouf, J.-F. & Simmons, A. The ECMWF operational implementation of four-dimensional variational assimilation. I: Experimental results with simplified physics. *Q. J. R. Meteorol. Soc.* **126**, 1143–1170 (2000).
36. Gauthier, P. & Thépaut, J.-N. Impact of the digital filter as a weak constraint in the preoperational 4DVAR assimilation system of Météo-France. *Mon. Weath. Rev.* **129**, 2089–2102 (2001).
37. Rawlins, F. *et al.* The Met Office global four-dimensional variational data assimilation scheme. *Q. J. R. Meteorol. Soc.* **133**, 347–362 (2007).
38. Kadowaki, T. A 4-dimensional variational assimilation system for the JMA Global Spectrum Model. *CAS/JSC WGN Res. Act. Atmos. Ocea. Modell.* **34**, 117–118 (2005).
39. Gauthier, P., Tanguay, M., Laroche, S., Pellerin, S. & Morneau, J. Extension of 3D-Var to 4D-Var: implementation of 4D-Var at the Meteorological Service of Canada. *Mon. Weath. Rev.* **135**, 2339–2354 (2007).
40. Xu, L., Rosmond, T. & Daley, R. Development of NAVDAS-AR: formulation and initial tests of the linear problem. *Tellus A* **57**, 546–559 (2005).
41. Saunders, R. W., Matricardi, M. & Brunel, P. An improved fast radiative transfer model for assimilation of satellite radiance observations. *Q. J. R. Meteorol. Soc.* **125**, 1407–1425 (1999).
42. Bauer, P., Moreau, E., Chevallier, F. & O’Keeffe, U. Multiple-scattering microwave radiative transfer for data assimilation applications. *Q. J. R. Meteorol. Soc.* **132**, 1259–1281 (2006).
43. Buehner, M. Ensemble-derived stationary and flow-dependent background-error covariances: evaluation in a quasi-operational NWP setting. *Q. J. R. Meteorol. Soc.* **131**, 1013–1043 (2006).
44. Chapnik, B., Desroziers, G., Rabier, F. & Talagrand, O. Diagnosis and tuning of observational error in a quasi-operational data assimilation setting. *Q. J. R. Meteorol. Soc.* **132**, 543–565 (2006).
45. Janisková, M. & Lopez, P. in *Data Assimilation for Atmospheric, Oceanic and Hydrological Applications* (eds Park, S. K. & Xu, L.) 251–286 (Springer, 2013).
46. Trenberth, K. E. & Fasullo, J. T. Climate extremes and climate change: the Russian heat wave and other climate extremes of 2010. *J. Geophys. Res.* **117**, D17103 (2012).
47. Hoskins, B. The potential for skill across the range of the seamless weather-climate prediction problem: a stimulus for our science. *Q. J. R. Meteorol. Soc.* **139**, 573–584 (2013).
48. Jung, T., Miller, M. J. & Palmer, T. N. Diagnosing the origin of extended-range forecast errors. *Mon. Weath. Rev.* **138**, 2434–2446 (2010).
49. Duc, L., Saito, K. & Seko, H. Spatial-temporal fractions verification for high-resolution ensemble forecasts. *Tellus A* **65**, 18171 (2013).
50. Rodwell, M. J., Richardson, D. S., Hewson, T. D. & Haiden, T. A new equitable score suitable for verifying precipitation in numerical weather prediction. *Q. J. R. Meteorol. Soc.* **136**, 1344–1363 (2010).
51. Johnson, C., Bowler, N. & I. On the reliability and calibration of ensemble forecasts. *Mon. Weath. Rev.* **137**, 1717–1720 (2009).
52. Martin, G. M. *et al.* Analysis and reduction of systematic errors through a seamless approach to modeling weather and climate. *J. Clim.* **23**, 5933–5957 (2010).
53. Rodwell, M. J. & Palmer, T. N. Using numerical weather prediction to assess climate models. *Q. J. R. Meteorol. Soc.* **133**, 129–146 (2007).
54. Dole, R. *et al.* The making of an extreme event: putting the pieces together. *Bull. Am. Meteorol. Soc.* **95**, 427–440 (2014).
55. Ferro, C. A. T. & Stephenson, D. B. Extremal Dependence Indices: improved verification measures for deterministic forecasts of rare binary events. *Weather Forecast.* **26**, 699–713 (2011).
56. Yu, H., Chen, P., Li, Q. & Tang, B. Current capability of operational numerical models in predicting tropical cyclone intensity in the western North Pacific. *Weather Forecast.* **38**, 353–367 (2013).
57. Molteni, F., Kucharski, F. & Corti, S. in *Predictability of Weather and Climate* (eds Palmer, T. & Hagedorn, R.) 365–390 (Cambridge Univ. Press, 2006).
58. Lin, H., Brunet, G. & Derome, J. An observed connection between the North Atlantic Oscillation and the Madden-Julian Oscillation. *J. Clim.* **22**, 364–380 (2009).
59. Barnston, A. G., Tippett, M. K., L’Heureux, M. L., Li, S. & DeWitt, D. G. Skill of Real-time seasonal ENSO model predictions during 2002–11: is our capability increasing? *Bull. Am. Meteorol. Soc.* **93**, 631–651 (2012).
60. Baldwin, M. P. & Dunkerton, T. J. Stratospheric harbingers of anomalous weather regimes. *Science* **244**, 581–584 (2001).
61. Scaife, A. A., Knight, J. R., Vallis, G. K. & Folland, C. K. A stratospheric influence on the winter NAO and North Atlantic surface climate. *Geophys. Res. Lett.* **32**, L18715 (2005).
62. Teng, H., Branstator, G. W., Wang, H., Meehl, G. A. & Washington, W. M. Probability of US heat waves affected by a subseasonal planetary wave pattern. *Nature Geosci.* **6**, 1056–1061 (2013).
63. Frank, W. M. & Roundy, P. E. The relationship between tropical waves and tropical cyclogenesis. *Mon. Weath. Rev.* **134**, 2397–2417 (2006).
64. Inness, P. M., Slingo, J. M., Woolnough, S. J., Neale, R. B. & Pope, V. D. Organisation of tropical convection in a GCM with varying vertical resolution; implications of the Madden Julian Oscillation. *Clim. Dyn.* **17**, 777–793 (2001).
65. Koster, R. D. *et al.* Contribution of land surface initialization to subseasonal forecast skill: first results from the GLACE-2 project. *Geophys. Res. Lett.* **37**, L02402 (2010).
66. Lin, H. & Brunet, G. The influence of the Madden-Julian oscillation on Canadian wintertime surface air temperature. *Mon. Weath. Rev.* **137**, 2250–2262 (2009).
67. Benoit, R. *et al.* Toward the use of coupled atmospheric and hydrologic models at regional scale. *Mon. Weath. Rev.* **128**, 1681–1706 (2000).

68. Zappa, M. *et al.* Propagation of uncertainty from observing systems and NWP into hydrological models: COST-731 Working Group 2. *Atmos. Sci. Lett.* **11**, 83–91 (2010).
 69. Lauritzen, P. H. & Thuburn, J. Evaluating advection/transport schemes using interrelated tracers, scatter plots and numerical mixing diagnostics. *Q. J. R. Meteorol. Soc.* **138**, 906–918 (2011).
 70. Zhang, Y., Bocquet, M., Mallet, V., Seigneur, C. & Baklanov, A. Real-time air quality forecasting, part II: State of the science, current research needs, and future prospects. *Atmos. Environ.* **60**, 656–676 (2012).
 71. Cloke, H. *et al.* Preface, hydrological ensemble prediction systems. *Hydrol. Processes* **27**, 1–4 (2013).
 72. Thorpe, A. J., Bauer, P., Magnusson, L. & Richardson, D. An evaluation of recent performance of ECMWF's forecasts. *ECMWF Newsl.* **137**, 15–18 (2013).
 73. Magnusson, L. & Källén, E. Factors influencing skill improvements in the ECMWF forecasting system. *Mon. Weath. Rev.* **141**, 3142–3153 (2013).
 74. Holloway, C. E., Woolnough, S. J. & Lister, G. M. S. The effects of explicit versus parameterized convection on the MJO in a large-domain high-resolution tropical case study. Part I: Characterization of large-scale organization and propagation. *J. Atmos. Sci.* **70**, 1342–1369 (2013).
 75. Hong, S. Y. J. D. Next-generation numerical weather prediction: bridging parameterization, explicit clouds, and large eddies. *Bull. Am. Meteorol. Soc.* **93**, ES6–ES9 (2012).
 76. Jakob, C. Accelerating progress in global atmospheric model development through improved parameterizations: challenges, opportunities, and strategies. *Bull. Am. Meteorol. Soc.* **91**, 869–875 (2010).
 77. Phillips, T. J. *et al.* Evaluating parameterizations in general circulation models: Climate simulation meets weather prediction. *Bull. Am. Meteorol. Soc.* **85**, 1903–1915 (2004).
 78. Shin, H. H. & Hong, S.-Y. Analysis of resolved and parameterized vertical transports in convective boundary layers at gray-zone resolutions. *J. Atmos. Sci.* **70**, 3248–3261 (2013).
 79. Watson, P. A. G., Christensen, H. M. & Palmer, T. N. Does the ECMWF IFS convection parameterization with stochastic physics correctly reproduce relationships between convection and the large-scale state? *J. Atmos. Sci.* **72**, 236–242 (2015).
 80. Randall, D., Khairoutdinov, M., Arakawa, A. & Grabowski, W. Breaking the cloud parameterization deadlock. *Bull. Am. Meteorol. Soc.* **84**, 1547–1564 (2003).
 81. MacLachlan, C. *et al.* Global Seasonal forecast system version 5 (GloSea5): a high-resolution seasonal forecast system. *Q. J. R. Meteorol. Soc.* **10.1002/qj.2396** (2014).
 82. Molteni, F. *et al.* The new ECMWF seasonal forecast system (system 4); ECMWF Technical Memoranda no. 656 (ECMWF, Reading, 2011).
 83. Gleckler, P. J. Surface energy balance errors in AGCMs: implications for ocean-atmosphere model coupling. *Geophys. Res. Lett.* **32**, L15708 (2005).
 84. Cardinali, C. in *Data Assimilation for Atmospheric, Oceanic and Hydrologic Applications* Vol. II (eds Park, S. K. & Xu, L.) 89–110 (Springer, 2013).
 85. Horanyi, A., Cardinali, C., Rennie, M. & Isaksen, I. The assimilation of horizontal line-of-sight wind information into the ECMWF data assimilation and forecasting system. Part I: The assessment of wind impact. *Q. J. R. Meteorol. Soc.* <http://dx.doi.org/10.1002/qj.2430> (2014).
 86. Lorenc, A. C., Bowler, N. E., Clayton, A. M., Pring, S. R. & Fairbairn, D. Comparison of hybrid-4DVar and hybrid-4DVar data assimilation methods for global NWP. *Mon. Weath. Rev.* **143**, 212–229 (2015).
 87. Evensen, G. *Data Assimilation: The Ensemble Kalman Filter* (Springer, 2007).
 88. Leeuwen, P. J. V. Efficient non-linear data assimilation in geophysical fluid dynamics. *Comput. Fluids*. <http://dx.doi.org/doi:10.1016/j.compfluid.2010.11.011> (2011).
 89. Brunet, G. *et al.* Collaboration of the weather and climate communities to advance subseasonal-to-seasonal prediction. *Bull. Am. Meteorol. Soc.* **91**, 1397–1406 (2010).
 90. Mollick, E. Establishing Moore's law. *IEEE Ann. Hist. Comput.* **28**, 62–75 (2006).
 91. Shalf, J., Dosanjh, S. & Morrison, J. in *High Performance Computing for Computational Science – VECPAR 2010* (eds Palma, J. M. L. M. *et al.*) 1–25 (Springer, 2011).
 92. Müller, E. H. & Scheichl, R. Massively parallel solvers for elliptic partial differential equations in numerical weather. *Q. J. R. Meteorol. Soc.* **140**, 2608–2624 (2014).
 93. Palem, K. Ten years of building broken chips: the physics and engineering of inexact computing. *ACM Trans. Embed. Comput. Syst. (TECS)* **12**, 87:1–87:23 (2013).
 94. Lawrence, B. *Data Handling on the Path to Exascale*; <http://old.ecmwf.int/newsevents/meetings/workshops/2014/Scalability/Presentations/pdfs/Lawrence.pdf> (presented at ECMWF Scalability Workshop, Reading, 2013).
 95. Collard, A. D., McNally, A. P., Hilton, F. I., Healy, S. B. & Atkinson, N. C. The use of principal component analysis for the assimilation of high-resolution infrared sounder observations for numerical weather prediction. *Q. J. R. Meteorol. Soc.* **136**, 2038–2050 (2010).
 96. Snik, F. *et al.* Mapping atmospheric aerosols with a citizen science network of smartphone spectropolarimeters. *Geophys. Res. Lett.* **41**, 7351–7358 (2014).
 97. Mahoney, W. P. III & O'Sullivan, J. M. Realizing the potential of vehicle-based observations. *Bull. Am. Meteorol. Soc.* **94**, 1007–1018 (2013).
 98. Hurrell, J. *et al.* A unified modeling approach to climate system prediction. *Bull. Am. Meteorol. Soc.* **90**, 1819–1832 (2009).
 99. World Meteorological Organization. *Seamless Prediction of the Earth System: From Minutes to Months*. WMO No. 1156 (World Meteorological Organization, Geneva, 2015).
 100. Simmons, A. J. & Hollingsworth, A. Some aspects of the improvement in skill of numerical weather prediction. *Q. J. R. Meteorol. Soc.* **128**, 647–677 (2002).
- Discusses the historical increase of the weather predictive skill of one day lead-time per decade linked to improvements in data assimilation, models and observing systems.**

Acknowledgements We thank C. Jakob and M. Miller for motivating us to write this paper. We are grateful to E. Källén, F. Rabier and A. Simmons for comments and to L. Magnusson for input and figures.

Author Contributions P.B. led the writing of the paper and conceived the ideas of Figs 5 and 6. All authors contributed to writing the manuscript and assembled the remaining figures.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to P.B. (peter.bauer@ecmwf.int)

The *C9orf72* repeat expansion disrupts nucleocytoplasmic transport

Ke Zhang^{1*}, Christopher J. Donnelly^{2*}, Aaron R. Haeusler³, Jonathan C. Grima^{2,4}, James B. Machamer¹, Peter Steinwald³, Elizabeth L. Daley², Sean J. Miller², Kathleen M. Cunningham¹, Svetlana Vidensky², Saksham Gupta¹, Michael A. Thomas², Ingie Hong⁴, Shu-Ling Chiu⁴, Richard L. Hagan⁴, Lyle W. Ostrow¹, Michael J. Matunis³, Jiou Wang³, Rita Sattler², Thomas E. Lloyd^{1,4§} & Jeffrey D. Rothstein^{2,4§}

The hexanucleotide repeat expansion (HRE) GGGGCC (G_4C_2) in *C9orf72* is the most common cause of amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD). Recent studies support an HRE RNA gain-of-function mechanism of neurotoxicity, and we previously identified protein interactors for the G_4C_2 RNA including RanGAP1. A candidate-based genetic screen in *Drosophila* expressing 30 G_4C_2 repeats identified RanGAP (*Drosophila* orthologue of human RanGAP1), a key regulator of nucleocytoplasmic transport, as a potent suppressor of neurodegeneration. Enhancing nuclear import or suppressing nuclear export of proteins also suppresses neurodegeneration. RanGAP physically interacts with HRE RNA and is mislocalized in HRE-expressing flies, neurons from *C9orf72* ALS patient-derived induced pluripotent stem cells (iPSC-derived neurons), and in *C9orf72* ALS patient brain tissue. Nuclear import is impaired as a result of HRE expression in the fly model and in *C9orf72* iPSC-derived neurons, and these deficits are rescued by small molecules and antisense oligonucleotides targeting the HRE G-quadruplexes. Nucleocytoplasmic transport defects may be a fundamental pathway for ALS and FTD that is amenable to pharmacotherapeutic intervention.

The G_4C_2 HRE in the *C9orf72* gene is found in as many as 40% of familial ALS and FTD cases, with additional reports in other neurodegenerative diseases^{1–3}. *C9orf72* HRE-induced cytotoxicity has been proposed to be caused through loss- and gain-of-function mechanisms that include: (1) transcribed sense GGGGCC_{exp} or antisense CCCCCG_{exp} RNAs that sequester proteins, thus altering their normal function²; or (2) the sense or antisense expanded RNAs are translated via repeat-associated non-AUG translation to form toxic dipeptide repeat proteins (DPRs)^{4–7}. We and others have demonstrated that HRE RNA forms hairpin and G-quadruplex structures that bind and sequester RNA binding proteins (RBPs)^{8,9}.

RanGAP suppresses HRE-mediated toxicity in *Drosophila*

We previously identified 19 proteins that exhibit high affinity to G_4C_2 relative to a G:C scrambled RNA, along with ~400 additional proteins that bind with a moderate affinity to G_4C_2 and/or bind both G_4C_2 and G:C scrambled RNA^{8,9}. To determine which of these candidate RBPs genetically modifies G_4C_2 -mediated neurodegeneration, we performed a screen in an established *Drosophila* model that expresses 30 G_4C_2 repeats ((G_4C_2)₃₀) in the fly eye¹⁰ (Supplementary Table 1). One of the strongest suppressors is a dominant, gain-of-function (GOF) allele of *RanGAP*, called *RanGAP^{SD}* (GOF), that functions similarly to overexpression of wild-type *RanGAP* (refs 11–13). As shown in Fig. 1a, 1-day-old flies expressing (G_4C_2)₃₀ display subtle ommatidial disorganization defects in the eye that worsen when aged for 15 days (Fig. 1b). However, flies expressing the same repeats in the heterozygous *RanGAP^{SD}* (GOF) mutant background or with *RanGAP* overexpression appear normal (Fig. 1a–c and Extended Data Fig. 1), indicating that *RanGAP* is a suppressor of G_4C_2 repeat toxicity.

As shown in Fig. 1a, b, wild-type fly eyes have seven photoreceptor neuron (PRN) rhabdomeres per ommatidium. In contrast, the PRNs expressing 30 G_4C_2 repeats show a loss of integrity and/or organization of rhabdomeres at day 15 (Fig. 1a, b, d), suggesting age-dependent degeneration. These phenotypes are rescued by either heterozygous *RanGAP^{SD}* (GOF) mutant or *RanGAP* overexpression. Conversely, knockdown of *RanGAP* by RNA interference (RNAi) significantly enhances the PRN defects (Fig. 1d and Extended Data Fig. 1a). Moreover, *RanGAP* knockdown-mediated enhancement of (G_4C_2)₃₀-mediated degeneration worsens with age, with an almost complete loss of rhabdomeres in aged flies which is not due to alterations in G_4C_2 mRNA level (Extended Data Fig. 1b). These data indicate that *RanGAP* is a potent suppressor of G_4C_2 -mediated neurodegeneration in the *Drosophila* eye.

To determine whether *RanGAP* also suppresses G_4C_2 -mediated toxicity in *Drosophila* motor neurons, we next analysed the effect of *RanGAP* overexpression on the locomotor function of adult flies. Neuronal expression of (G_4C_2)₃₀ throughout adulthood causes flight defects in 15-day-old flies that are rescued with simultaneous overexpression of *RanGAP* (Extended Data Fig. 1c, d). Interestingly, when expressed in motor neurons throughout larval development using *OK371-GAL4*, (G_4C_2)₃₀ causes severe neuromuscular junction (NMJ) defects including an ~50% reduction in active zone number and impaired evoked neurotransmitter release not rescued by *RanGAP* overexpression (Extended Data Fig. 2). Together, these data suggest that *RanGAP* suppresses G_4C_2 -mediated neurodegeneration during adulthood, whereas (G_4C_2)₃₀ expression during development causes neurotoxicity that is independent of *RanGAP*.

¹Department of Neurology, School of Medicine, Johns Hopkins University, Maryland 21205, USA. ²Brain Science Institute, School of Medicine, Johns Hopkins University, Maryland 21205, USA.

³Biochemistry and Molecular Biology, Bloomberg School of Public Health, Johns Hopkins University, Maryland 21205, USA. ⁴Department of Neuroscience, School of Medicine, Johns Hopkins University, Maryland 21205, USA.

*These authors contributed equally to this work.

§These authors jointly supervised this work.

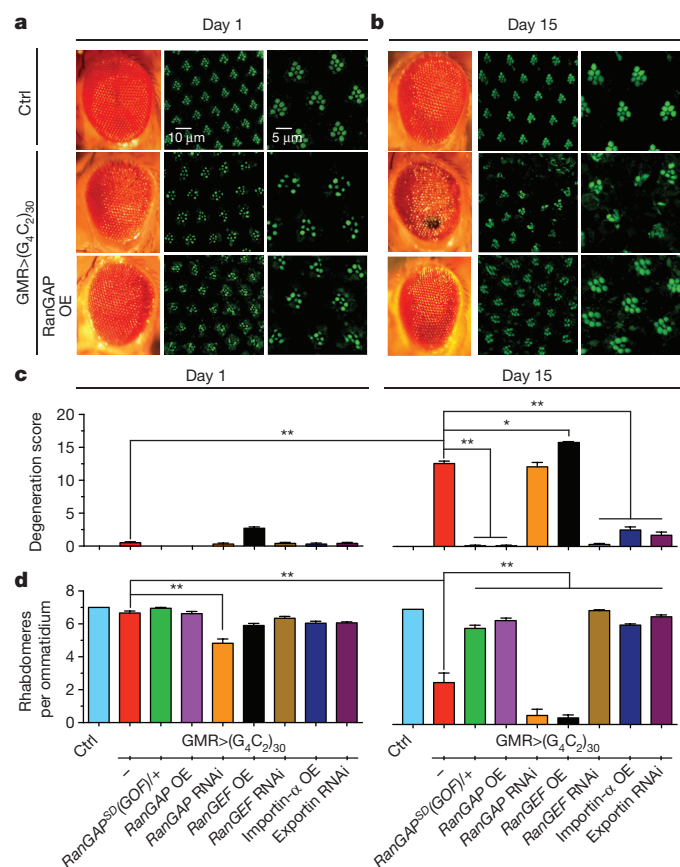


Figure 1 | Genetic interaction between G_4C_2 repeats and nucleocytoplasmic transport machinery. **a, b**, External eye morphology of 1-day-old (**a**, left panels) and 15-day-old (**b**, left panels) flies. Phalloidin staining of the retina of 1-day-old (**a**, middle panels, magnified in right panels) and 15-day-old (**b**, middle panels, magnified in right panels) flies. Wild-type control (top row), flies expressing 30 G_4C_2 repeats (middle row), and flies expressing 30 G_4C_2 repeats and overexpressing RanGAP (bottom row) are shown. Genotypes: top row, *GMR-GAL4/+*; middle row, *GMR-GAL4, UAS-(G₄C₂)_{30/+}*; bottom row, *GMR-GAL4, UAS-(G₄C₂)_{30/+}; UAS-RanGAP/+*. **c, d**, Quantification of external morphology (**c**) and rhabdomere number (**d**). **P* < 0.05; ***P* < 0.01.

Nucleocytoplasmic transport modulates G₄C₂ toxicity

RanGAP functions in the cytoplasm to stimulate Ran GTPase (hereafter referred to as Ran) to hydrolyse GTP to GDP, a process required for efficient nucleocytoplasmic transport^{12,14,15}. Proteins larger than 40 kDa require active transport to cross the nuclear pore complex (NPC), in which their nuclear localization sequence (NLS) and/or nuclear export signal (NES) are recognized by carrier protein importins and/or exportins, respectively^{16–18}. In the nucleus, exportins bind Ran•GTP and cargo proteins for export. Nuclear Ran guanine nucleotide exchange factor (RanGEF) converts Ran•GDP back to Ran•GTP¹⁹.

As shown in Fig. 1c, d and Extended Data Fig. 1a, b, overexpression of RanGEF enhances G_4C_2 -repeat-mediated degeneration, resulting in large necrotic patches and severe rhabdomyere degeneration. In contrast, knockdown of RanGEF rescues both of these phenotypes. These data suggest that RanGEF has an opposite role when compared with RanGAP in G_4C_2 -mediated neurodegeneration, consistent with their opposing biochemical functions. Overexpression of importin- α or knockdown of exportin rescues these phenotypes. Thus, genetically enhancing nuclear import or inhibiting export of NLS/NES-containing proteins suppresses G_4C_2 -mediated neurodegeneration.

Expression of arginine-containing DPRs in *Drosophila* causes severe toxicity and poly-glycine-arginine (GR) DPRs are detected in flies expressing 36 G₄C₂ repeats under the control of heat-shock-inducible GAL4 (*hs-GAL4*)⁷. Although we are also able to detect polyGR DPRs in

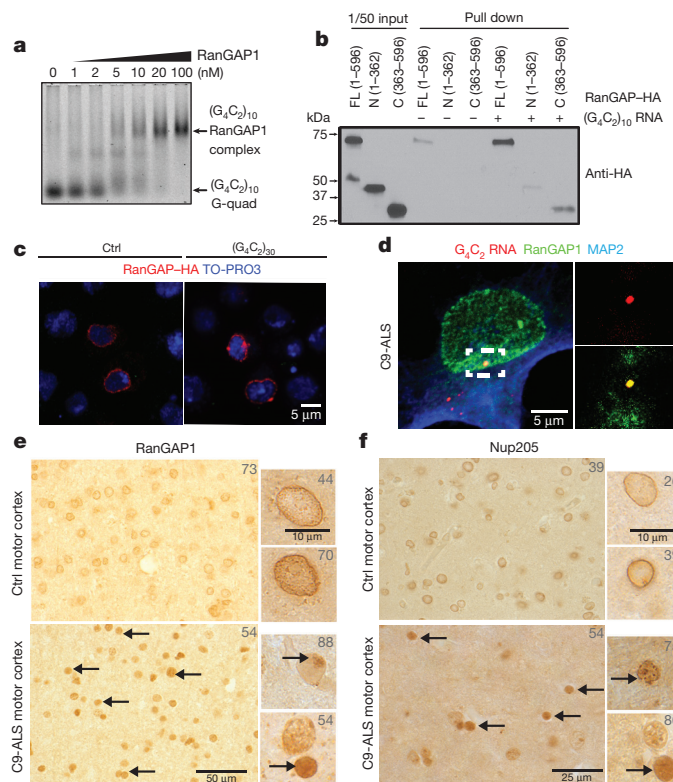


Figure 2 | RanGAP binds to G₄C₂ repeats and is mislocalized along with NPC components. **a**, EMSA of human RanGAP1 and repeat RNA in the G-quadruplex conformation. **b**, RanGAP-HA pull down in the absence (lanes 4–6) or presence (lanes 7–9) of biotinylated G₄C₂ RNA repeats, immunoblotted with a HA antibody. Lanes 1–3: 1/50 input. FL, full length. **c**, Wild-type control (left) and G₄C₂ HRE (right) S2 cells expressing RanGAP-HA co-stained with an antibody against HA (red) and TO-PRO3 (blue). **d**, RanGAP1 co-localization with G₄C₂ RNA foci (dotted box: projected view, high magnification: ~0.3 μm single plane) in a C9-ALS iPSC neuron in confocal single plane image. **e**, RanGAP1 immunostaining in non-neurological disease control and C9orf72 ALS motor cortex showing intense nuclear localization (arrows) and aberrant nuclear aggregates (individual patient identifier in upper right corner, Supplementary Table 2). **f**, Abnormal nuclear localization of Nup205 in C9orf72 human motor cortex cells.

hs-GAL4, *UAS-(G₄C₂)₃₀* flies when heat shocked, we are not able to detect polyGR or polyGP DPRs when (G₄C₂)₃₀ is expressed in the eye with *GMR-GAL4* at the time of PRN degeneration or in adult neurons with *elavGS* (Extended Data Fig. 3a, b). Nonetheless, we cannot exclude the possibility that DPRs are expressed at undetectable levels and contribute to degeneration in the eye.

G₄C₂ repeats bind RanGAP and cause NPC pathology

To determine the relative affinity of RanGAP for G₄C₂ RNA, we performed an electrophoretic gel mobility shift assay (EMSA) with (G₄C₂)₁₀ RNA and recombinant human RanGAP1 (Extended Data Fig. 4a). The sense (G₄C₂)₁₀ RNA G-quadruplex shows a concentration- and length-dependent shift of free RNA to a lower mobility RNA–RanGAP1 complex with increasing concentrations of RanGAP1 (Fig. 2a and Extended Data Fig. 4b–d). Additionally, RanGAP1 demonstrates a higher binding affinity to the sense strand G-quadruplex compared to hairpins (Extended Data Fig. 4b–d), and very little interaction was observed between RanGAP1 and (CUG)₂₀. Furthermore, the RanGAP1–(G₄C₂)₁₀ complex is resistant to antisense oligonucleotides against the G₄C₂ repeat and nonspecific RNA competitor even at 1,000-fold RNA molar excess (Extended Data Fig. 4e). These *in vitro* results indicate that RanGAP1 preferentially binds the sense RNA G-quadruplex from the C9orf72 HRE.

To confirm that RanGAP interacts with G_4C_2 RNA in cells, we expressed carboxy-terminal haemagglutinin (HA)-tagged *Drosophila* RanGAP protein in S2 cells and performed RNA pull down. As shown in Fig. 2b, western blot analysis demonstrates that the full-length and C-terminal domain of RanGAP physically interact with G_4C_2 -repeat RNA in *Drosophila* cells. Both endogenous and transfected RanGAP–HA uniformly surrounds the nuclei of control cells (Fig. 2c and Extended Data Fig. 5a, b). In contrast, expression of $(G_4C_2)_{30}$ leads to formation of large RanGAP perinuclear puncta, which is not due to induction of apoptosis (Extended Data Fig. 5a).

In parallel with studies in *Drosophila*, we investigated RanGAP1 in iPSC-derived neurons (hereafter referred to as iPSC neurons) derived from multiple *C9orf72* ALS (C9-ALS) patients (Extended Data Fig. 6 and Supplementary Tables 3 and 4). Our iPSC cultures comprise 30–40% motor neurons (Islet-1⁺) and are predominantly excitatory neurons (vGLUT⁺) and express additional motor neuron markers, neural-specific cytoskeletal proteins and synaptic proteins (Supplementary Table 3 and Extended Data Fig. 6b, c). Consistent with our observations in S2 cells, iPSC neurons from C9-ALS patients variably exhibit RanGAP1 puncta (Extended Data Fig. 5c), and RanGAP1 can co-localize with G_4C_2 RNA (Fig. 2d). Notably, to determine whether RanGAP1 mislocalization occurs in human disease, we analysed brain tissue of C9-ALS patients. Cells in C9-ALS motor cortex commonly exhibit mislocalized, discontinuous and large punctate RanGAP1 signals compared to smooth perinuclear staining observed in controls (Fig. 2e, Extended Data Fig. 7a–d and Supplementary Table 2). Similar pathology was not readily observed in C9-ALS cerebellar cortex (Extended Data Fig. 7e). Perinuclear cytoplasmic RanGAP1 puncta occasionally co-localize with ubiquitin (Extended Data Fig. 5e). We next asked whether RanGAP1 puncta contain other protein components of the NPC, and therefore stained for nucleoporin 205 (Nup205), an extremely long-lived NPC scaffold protein^{20,21}. We found that RanGAP1 and Nup205 co-localize and are predominantly perinuclear in control iPSC neurons and brain tissue (Fig. 2f). Interestingly, Nup205 co-localizes with some RanGAP1 aggregates in C9-ALS iPSC neurons (Extended Data Fig. 5d). Consistent with this observation, Nup205 and Nup107 exhibit similar motor cortex pathology as RanGAP1 in multiple C9-ALS patients (Fig. 2f, Extended Data Fig. 7f and Supplementary Table 2). These data suggest that RanGAP1 and additional components of the NPC are disrupted in C9-ALS patients.

The Ran gradient is disrupted by the *C9orf72* HRE

We then tested whether sequestration of RanGAP by the G_4C_2 RNA leads to its loss of function. Most Ran protein is imported into the nucleus, a process that requires its binding to GDP, but not GTP^{22–24}. Hence, defects in RanGAP1 activity might affect the nuclear–cytoplasmic (N/C) distribution of Ran. Indeed, we observed a significant reduction in the N/C ratio of Ran in S2 cells expressing $(G_4C_2)_{30}$ (Fig. 3a), suggesting that RanGAP function is impaired.

Next, we quantified nuclear and cytoplasmic Ran in C9-ALS iPSC neurons via immunofluorescence (Fig. 3b–d and Extended Data Fig. 8b, c). We observed a significant reduction in the N/C ratio of endogenous Ran in the C9-ALS lines tested in mature MAP2⁺ iPSC neurons at 50–70 days *in vitro* (DIV) (Fig. 3c, d). Ran gradient abnormalities were also detected in mature ChAT⁺ neurons within the same cultures (Fig. 3e and Extended Data Fig. 8a). Overexpression of a functional Ran–GFP fusion protein in C9-ALS iPSC neurons also showed reduced N/C Ran gradients in C9-ALS iPSC neurons (Extended Data Fig. 8b, c).

RanGAP1–GFP overexpression in C9-ALS iPSC neurons rescued the disrupted N/C Ran gradient to control levels (Extended Data Fig. 8d), demonstrating that altered RanGAP1 function contributes to the disrupted N/C Ran ratio. The disrupted Ran gradient is not due to apoptosis, since treatment of control iPSC neurons with tunicamycin does not alter the N/C Ran ratio despite elevating activated caspase

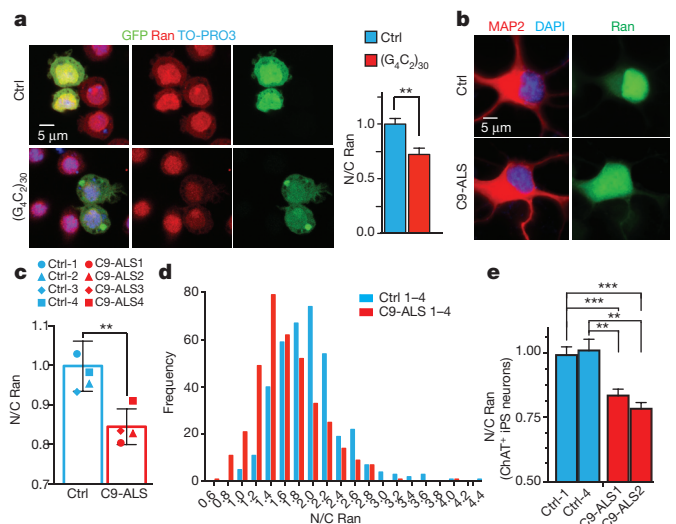


Figure 3 | *C9orf72* HRE disrupts the nuclear/cytoplasmic Ran gradient.

a, S2 cells co-transfected with GFP and $(G_4C_2)_{30}$ (bottom row) or control (top row) and stained with a Ran antibody (red) and TO-PRO3 (blue). Error bars indicate s.e.m. **b**, iPSC neurons from control and C9-ALS patients showing mislocalization of Ran to the cytoplasm in C9-ALS iPSC neurons. **c**, Quantification of N/C Ran gradient in neurons from four control and four C9-ALS iPS lines when normalized to control. N/C Ran ratio is reduced in C9-ALS neurons. Each symbol represents mean of up to 228 neurons per line (see Supplementary Table 4). Bar indicates mean N/C Ran of four control or C9-ALS lines; error bars indicate s.e.m. **d**, N/C Ran histogram shows higher frequency of lower N/C ratios in four C9-ALS lines as compared to the four control lines. N/C ratios are presented as raw values. **e**, C9-ALS ChAT⁺ neurons show similar reduction of N/C Ran. N/C Ran is normalized to controls and up to 60 neurons were tested per line (see Supplementary Table 4) (** $P < 0.01$, *** $P < 0.0001$). Error bars indicate s.e.m.

3 levels (Extended Data Fig. 8e), and RanGAP1 and Ran mislocalization were not observed in iPSC astrocytes derived from C9-ALS patients (Extended Data Figs 6d and 8f–i). Taken together, our fly and human iPSC data indicate that the G_4C_2 HRE impairs neuronal RanGAP1 function, resulting in higher levels of cytoplasmic Ran protein.

The *C9orf72* HRE inhibits import of nuclear proteins

To determine whether the HRE significantly impairs nuclear import, we overexpressed a GFP protein tagged with both a classical NLS and a NES (NLS–NES–GFP)¹² in the *Drosophila* salivary gland where the cytoplasm and nucleus are large and distinct. NLS–NES–GFP is localized to both nuclei and cytoplasm of wild-type salivary gland cells (Fig. 4a). However, in cells expressing $(G_4C_2)_{30}$, the N/C ratio of NLS–NES–GFP is severely reduced (Fig. 4a and Extended Data Fig. 9a), suggesting that nuclear import is inhibited and/or that nuclear export is enhanced.

Next, we expressed a GFP protein tagged with an NLS and a mutated NES, which severely impairs its export activity (Δ NES)¹². In control cells, NLS– Δ NES–GFP localizes primarily to the nucleus (Fig. 4a), whereas in cells expressing $(G_4C_2)_{30}$ it localizes predominantly to the cytoplasm (Fig. 4a and Extended Data Fig. 9a), supporting an impairment of nuclear import in these cells. Using immunoblot analysis, we confirmed that the levels of GFP protein are similar in control and $(G_4C_2)_{30}$ -expressing salivary glands (Extended Data Fig. 9b). We also detected cytoplasmic mislocalization of NLS–NES–GFP and NLS– Δ NES–GFP in glutamatergic neurons of the ventral nerve cord in $(G_4C_2)_{30}$ -expressing flies (Extended Data Fig. 9d), indicating that the G_4C_2 HRE also affects nucleocytoplasmic transport in *Drosophila* motor neurons. Therefore, expression of the G_4C_2 HRE decreases nuclear import in *Drosophila* cells *in vivo*.

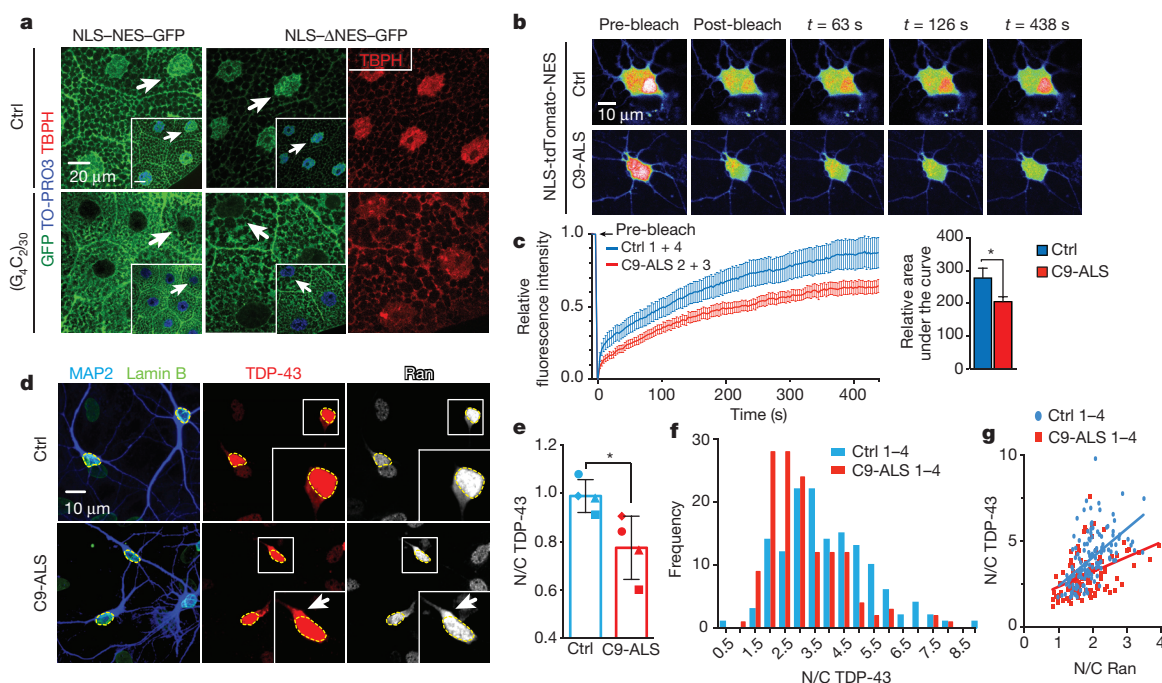


Figure 4 | *C9orf72* HRE causes nucleocytoplasmic transport defects.

a, Salivary glands expressing NLS-NES-GFP or NLS-ΔNES-GFP were co-stained for GFP, TBPH (red) and nuclei (blue, insets). Scale bars, 20 μ m. **b**, Representative images of NLS-tdTomato-NES FRAP analysis in control and C9-ALS iPS neurons (control, $n = 34$; C9-ALS, $n = 29$). **c**, Quantification of nuclear recovery (FRAP) of two C9-ALS and control iPS lines. Error bars indicate s.e.m. **d**, Representative images of control and C9-ALS iPS neurons. Arrows indicate higher cytoplasmic Ran and TDP-43 signals. **e**, Quantification

of mean N/C ratio of TDP-43 of four control and four C9-ALS lines when normalized to controls. Each symbol represents up to 49 neurons per line (see Supplementary Table 4). Error bars indicate s.d. **f**, Histogram shows higher frequency of lower N/C TDP-43 ratio. N/C ratios are presented as raw values. **g**, N/C TDP-43 directly correlates with N/C Ran ratio across all lines tested. N/C TDP-43 versus N/C Ran: control, $P < 0.0001$, $r^2 = 0.2980$; C9-ALS, $P < 0.0001$, $r^2 = 0.1657$. * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$.

We next investigated the effects of nuclear import deficits on candidate nuclear NLS- or NES-containing proteins. TDP-43 (TBPH in *Drosophila*), a predominantly nuclear protein, contains both a classical NLS and NES, and it is depleted from the nucleus of some CNS neurons and glia in most ALS patients and ~45% of FTD patients²⁵. Therefore, we hypothesize that its nuclear localization will be affected if nuclear import is disrupted. Indeed, loss of nuclear Ran correlates with depletion of nuclear TDP-43 in an FTD mouse model²⁶. As shown in Fig. 4a and Extended Data Fig. 9c, the N/C ratio of endogenous TBPH is significantly reduced in $(G_4C_2)_{30}$ -expressing salivary gland cells.

To validate our observations in human neurons, we investigated nucleocytoplasmic transport in iPS neurons by expressing a tdTomato reporter with a classical NLS and NES and performing fluorescence recovery after photobleaching (FRAP) of neuronal nuclei²⁷. We observed reduced nuclear recovery of NLS-tdTomato-NES in C9-ALS iPS neurons when compared with control lines (Fig. 4b, c). This defect was associated with disruption of TDP-43 localization, as C9-ALS iPS neurons exhibit variable, but significantly reduced, N/C ratios for TDP-43 (Fig. 4d–f). N/C ratios of Ran and TDP-43 correlate in control and C9-ALS iPS neurons (Fig. 4g), consistent with previous findings that the nuclear import of TDP-43 is Ran dependent²⁸. These data indicate that the C9-ALS HRE leads to impaired nuclear import of proteins that contain a classical NLS.

Rescue of HRE-mediated neurodegeneration

To determine whether antisense oligonucleotides targeting the *C9orf72* RNA rescue the disrupted N/C Ran ratio observed in C9-ALS iPS neurons, we treated these cells with C9 sense or scrambled antisense oligonucleotides used previously^{8,29,30}. The sense strand antisense oligonucleotide treatment reduced RNA foci in C9-ALS iPS neurons (Extended Data Fig. 8j) and fully rescued the disrupted N/C Ran gradient (Fig. 5a), suggesting that the nucleocytoplasmic

transport deficits are due to *C9orf72* sense-strand toxicity. Notably, when C9-ALS iPS neurons were treated with these antisense oligonucleotides, both the N/C Ran and TDP-43 gradients were increased (Extended Data Fig. 8k). The antisense oligonucleotide also suppressed nuclear import defects caused by G_4C_2 repeats *in vivo*. *Drosophila* larvae co-expressing $(G_4C_2)_{30}$ and NLS-ΔNES-GFP were raised on food supplemented with an antisense oligonucleotide throughout larval stages, mitigating nuclear mislocalization of NLS-ΔNES-GFP in salivary glands (Fig. 5b).

RanGAP1 binds the G_4C_2 RNA G-quadruplex *in vitro* (Fig. 2a). Therefore, we then tested whether this interaction can be perturbed by a porphyrin compound, TMPyP4, that destabilizes RNA G-quadruplex tertiary structures³¹. TMPyP4 reduces the affinity of RanGAP1 for the $(G_4C_2)_{10}$ G-quadruplex in a dose-dependent manner (Fig. 5c). TMPyP4 also rescues nuclear import defects in the fly model in a dose-dependent manner (Fig. 5d). Thus, inhibition of the G_4C_2 G-quadruplex structure significantly suppresses HRE-mediated nuclear import deficits. Interestingly, these phenotypes are also suppressed using an exportin 1 inhibitor, KPT-276 (Fig. 5e)³², suggesting that inhibiting nuclear export may compensate for disrupted import. Notably, antisense oligonucleotide, KPT-276, or TMPyP4 treatments all significantly suppress G_4C_2 -mediated neurodegeneration in the eye (Fig. 5f). Hence, our data suggest that modulation of nucleocytoplasmic transport presents a potential therapeutic strategy for neurodegenerative diseases characterized by the *C9orf72* HRE.

Discussion

Our data demonstrate that the G_4C_2 repeat expansion disrupts nucleocytoplasmic transport in a fly model and in human cells (Extended Data Fig. 10). While our data suggest that RanGAP is a key target of the G_4C_2 repeat expansion, other members of the NPC may also interact directly or indirectly with G_4C_2 . Several human genetic studies have implicated nuclear transport deficits as the cause

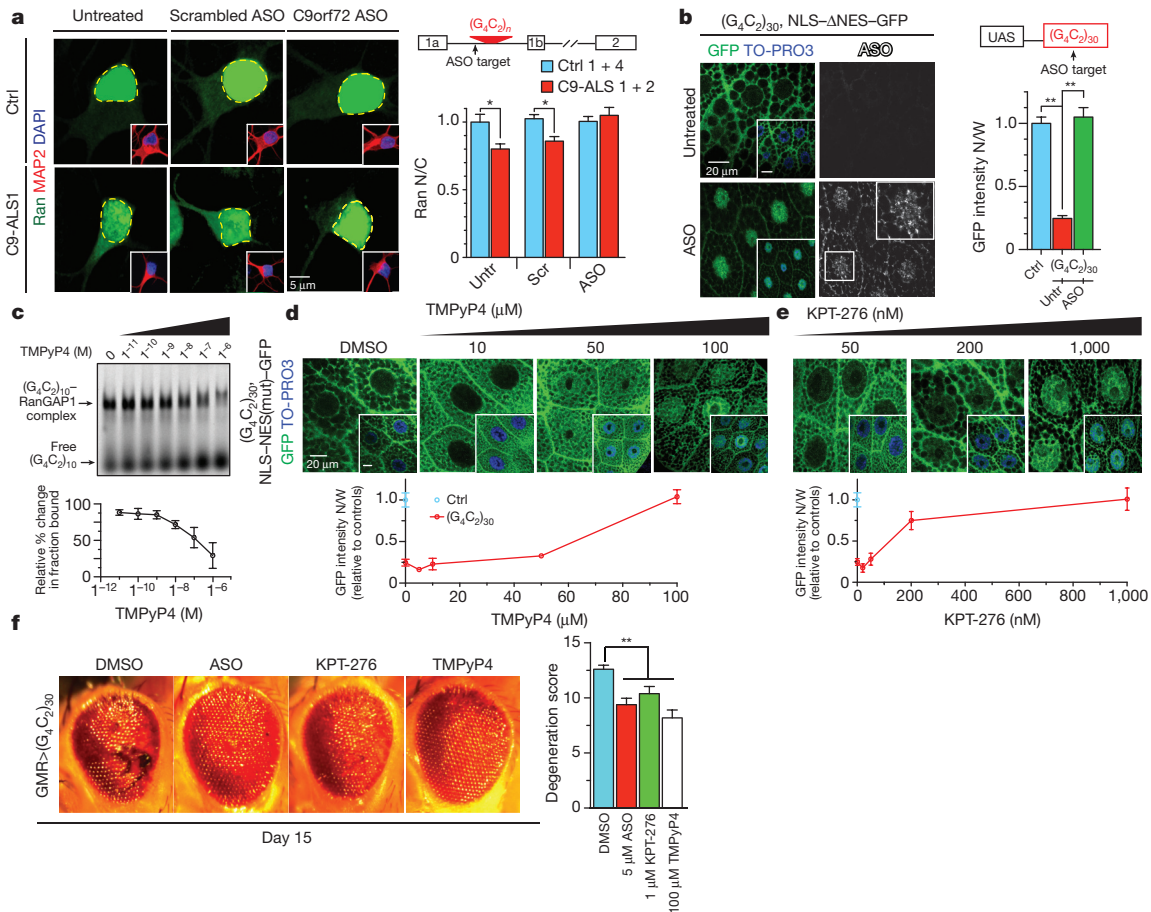


Figure 5 | Pharmacological rescue of nucleocytoplasmic transport defects. **a**, Neuronal N/C Ran ratio in control and two C9-ALS iPSC lines show increased cytoplasmic Ran levels in untreated and scrambled antisense-oligonucleotide-treated C9-ALS iPSC neurons ($n = 50$ neurons per line; see Supplementary Table 4). **b**, Salivary glands of larvae expressing G_4C_2 HRE and NLS-ΔNES-GFP were untreated (top) or treated with 5 μM antisense oligonucleotide (ASO, bottom) and co-stained for GFP (green), TO-PRO3 (blue) and antisense oligonucleotide (white). N, nuclear; W, whole cell. **c**, EMSA of RanGAP1 and

repeat RNA in the presence of TMPyP4 (top panel) and relative change in fraction bound (bottom panel). **d**, **e**, Salivary glands of larvae expressing G_4C_2 HRE and NLS-ΔNES-GFP were treated with different concentrations of TMPyP4 (**d**) or KPT-276 (**e**) versus vehicle control and co-stained for GFP (green) and TO-PRO3 (blue). Scale bars, 20 μm. **f**, The effects of antisense oligonucleotide, KPT-276 and TMPyP4 on the external morphology of eyes expressing G_4C_2 repeats. * $P < 0.05$; ** $P < 0.01$. All error bars indicate s.e.m.

of a rare fetal motor neuron disease and infrequent cases of ALS, including studies on the role of the nucleoporin GLE1 implicated in mRNA export^{33,34}. In addition, irregularities of the nuclear membrane and distribution of nuclear pore proteins were recently noted in sporadic ALS tissue³⁵. An accompanying paper (ref. 36) independently identified additional components of the NPC and nucleocytoplasmic trafficking pathways as dominant modifiers of G_4C_2 HRE toxicity in another C9-ALS fly model. Importantly, the observed NPC and nucleocytoplasmic trafficking defects in both iPSC-cell-derived neurons and motor neurons in our study are relevant to both ALS and FTD. Taken together, these studies suggest that products of the *C9orf72* HRE disrupt nucleocytoplasmic transport at the NPC and are a fundamental mechanism for inducing cellular injury in ALS and FTD. These defects may account for the nuclear depletion and cytoplasmic accumulation of TDP-43 widely seen in C9-ALS and FTD.

Although our data only demonstrate a role for disruption of nuclear import in C9-ALS pathogenesis, the robust nuclear pore pathology that we detect suggests that both nuclear import and export may be affected. It is enticing to speculate that NPC dysfunction leads to age-related neurodegeneration, since many of the NPC components, including Nup205, are extremely long-lived²⁰, and NPC integrity is lost during normal ageing³⁷.

The sense strand appears to be the cause of the described nucleocytoplasmic trafficking deficits in our human and fly model systems, as small molecules targeting the sense RNA suppress the nuclear import phenotypes, and neurodegeneration is caused by expression of G_4C_2 repeat RNA in *C9orf72* iPSC neurons or *Drosophila*. While we cannot exclude DPRs as a contributor to nucleocytoplasmic trafficking defects, our data in multiple model systems are most consistent with an RNA-mediated mechanism. Future studies will be required to determine the contribution of RanGAP disruption in C9-ALS pathogenesis compared with other pathogenic mechanisms implicated in C9-ALS such as nucleolar stress, which could act independently or in conjunction with nucleocytoplasmic transport disruption⁹.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 21 December 2014; accepted 24 July 2015.

Published online 26 August 2015.

1. Majounie, E. *et al.* Frequency of the *C9orf72* hexanucleotide repeat expansion in patients with amyotrophic lateral sclerosis and frontotemporal dementia: a cross-sectional study. *Lancet Neurol.* **11**, 323–330 (2012).
2. DeJesus-Hernandez, M. *et al.* Expanded GGGGCC hexanucleotide repeat in noncoding region of *C9ORF72* causes chromosome 9p-linked FTD and ALS. *Neuron* **72**, 245–256 (2011).

3. Renton, A. E. *et al.* A hexanucleotide repeat expansion in C9orf72 is the cause of chromosome 9p21-linked ALS-FTD. *Neuron* **72**, 257–268 (2011).
4. Ash, P. E. *et al.* Unconventional translation of C9orf72 GGGGCC expansion generates insoluble polypeptides specific to c9FTD/ALS. *Neuron* **77**, 639–646 (2013).
5. Mori, K. *et al.* The C9orf72 GGGGCC repeat is translated into aggregating dipeptide-repeat proteins in FTD/ALS. *Science* **339**, 1335–1338 (2013).
6. Kwon, I. *et al.* Poly-dipeptides encoded by the C9orf72 repeats bind nucleoli, impede RNA biogenesis, and kill cells. *Science* **345**, 1139–1145 (2014).
7. Mizielińska, S. *et al.* C9orf72 repeat expansions cause neurodegeneration in *Drosophila* through arginine-rich proteins. *Science* **345**, 1192–1194 (2014).
8. Donnelly, C. J. *et al.* RNA toxicity from the ALS/FTD C9orf72 expansion is mitigated by antisense intervention. *Neuron* **80**, 415–428 (2013).
9. Haessler, A. R. *et al.* C9orf72 nucleotide repeat structures initiate molecular cascades of disease. *Nature* **507**, 195–200 (2014).
10. Xu, Z. *et al.* Expanded GGGGCC repeat RNA associated with amyotrophic lateral sclerosis and frontotemporal dementia causes neurodegeneration. *Proc. Natl Acad. Sci. USA* **110**, 7778–7783 (2013).
11. Kusano, A., Staber, C. & Ganetzky, B. Segregation distortion induced by wild-type RanGAP in *Drosophila*. *Proc. Natl Acad. Sci. USA* **99**, 6866–6870 (2002).
12. Kusano, A., Staber, C. & Ganetzky, B. Nuclear mislocalization of enzymatically active RanGAP causes segregation distortion in *Drosophila*. *Dev. Cell* **1**, 351–361 (2001).
13. Merrill, C., Bayraktaroglu, L., Kusano, A. & Ganetzky, B. Truncated RanGAP encoded by the Segregation Distorter locus of *Drosophila*. *Science* **283**, 1742–1745 (1999).
14. Bischoff, F. R., Krebber, H., Kempf, T., Hermes, I. & Ponstingl, H. Human RanGTPase-activating protein RanGAP1 is a homologue of yeast Rna1p involved in mRNA processing and transport. *Proc. Natl Acad. Sci. USA* **92**, 1749–1753 (1995).
15. Bischoff, F. R., Klebe, C., Kretschmer, J., Wittinghofer, A. & Ponstingl, H. RanGAP1 induces GTPase activity of nuclear Ras-related Ran. *Proc. Natl Acad. Sci. USA* **91**, 2587–2591 (1994).
16. Steggerda, S. M. & Paschal, B. M. Regulation of nuclear import and export by the GTPase Ran. *Int. Rev. Cytol.* **217**, 41–91 (2002).
17. Mor, A., White, M. A. & Fontoura, B. M. Nuclear trafficking in health and disease. *Curr. Opin. Cell Biol.* **28**, 28–35 (2014).
18. Corbett, A. H. & Krebber, H. Hot trends erupting in the nuclear transport field. Workshop on mechanisms of nuclear transport. *EMBO Rep.* **5**, 453–458 (2004).
19. Frasch, M. The maternally expressed *Drosophila* gene encoding the chromatin-binding protein B1 is a homolog of the vertebrate gene Regulator of Chromatin Condensation, RCC1. *EMBO J.* **10**, 1225–1236 (1991).
20. Toyama, B. H. *et al.* Identification of long-lived proteins reveals exceptional stability of essential cellular structures. *Cell* **154**, 971–982 (2013).
21. Galy, V., Mattaj, J. W. & Askjaer, P. *Caenorhabditis elegans* nucleoporins Nup93 and Nup205 determine the limit of nuclear pore complex size exclusion *in vivo*. *Mol. Biol. Cell* **14**, 5104–5115 (2003).
22. Ribbeck, K., Lipowsky, G., Kent, H. M., Stewart, M. & Gorlich, D. NTF2 mediates nuclear import of Ran. *EMBO J.* **17**, 6587–6598 (1998).
23. Smith, A. E., Slepchenko, B. M., Schaff, J. C., Loew, L. M. & Macara, I. G. Systems analysis of Ran transport. *Science* **295**, 488–491 (2002).
24. Smith, A., Brownawell, A. & Macara, I. G. Nuclear import of Ran is mediated by the transport factor NTF2. *Curr. Biol.* **8**, 1403–1406 (1998).
25. Neumann, M. *et al.* Ubiquitinated TDP-43 in frontotemporal lobar degeneration and amyotrophic lateral sclerosis. *Science* **314**, 130–133 (2006).
26. Ward, M. E. *et al.* Early retinal neurodegeneration and impaired Ran-mediated nuclear import of TDP-43 in progranulin-deficient FTD. *J. Exp. Med.* **211**, 1937–1945 (2014).
27. D'Angelo, M. A., Gomez-Cavazos, J. S., Mei, A., Lackner, D. H. & Hetzer, M. W. A change in nuclear pore complex composition regulates cell differentiation. *Dev. Cell* **22**, 446–458 (2012).
28. Nishimura, A. L. *et al.* Nuclear import impairment causes cytoplasmic trans-activation response DNA-binding protein accumulation and is associated with frontotemporal lobar degeneration. *Brain* **133**, 1763–1771 (2010).
29. Sareen, D. *et al.* Targeting RNA foci in iPSC-derived motor neurons from ALS patients with a C9orf72 repeat expansion. *Sci. Transl. Med.* **5**, 208ra149 (2013).
30. Lagier-Tourenne, C. *et al.* Targeted degradation of sense and antisense C9orf72 RNA foci as therapy for ALS and frontotemporal degeneration. *Proc. Natl Acad. Sci. USA* **110**, E4530–E4539 (2013).
31. Zamiri, B., Reddy, K., Macgregor, R. B. Jr & Pearson, C. E. TMPyP4 porphyrin distorts RNA G-quadruplex structures of the disease-associated r(GGGGCC)_n repeat of the C9orf72 gene and blocks interaction of RNA-binding proteins. *J. Biol. Chem.* **289**, 4653–4659 (2014).
32. Schmidt, J. *et al.* Genome-wide studies in multiple myeloma identify XPO1/CRM1 as a critical target validated using the selective nuclear export inhibitor KPT-276. *Leukemia* **27**, 2357–2365 (2013).
33. Nousiainen, H. O. *et al.* Mutations in mRNA export mediator GLE1 result in a fetal motoneuron disease. *Nature Genet.* **40**, 155–157 (2008).
34. Kaneh, H. M. *et al.* Deleterious mutations in the essential mRNA metabolism factor, hGle1, in amyotrophic lateral sclerosis. *Hum. Mol. Genet.* **24**, 1363–1373 (2015).
35. Kinoshita, Y. *et al.* Nuclear contour irregularity and abnormal transporter protein distribution in anterior horn cells in amyotrophic lateral sclerosis. *J. Neuropathol. Exp. Neurol.* **68**, 1184–1192 (2009).
36. Freibaum, B. D. *et al.* GGGGCC repeat expansion in C9orf72 compromises nucleocytoplasmic transport. *Nature*. <http://dx.doi.org/10.1038/nature14974> (2015).
37. D'Angelo, M. A., Raices, M., Panowski, S. H. & Hetzer, M. W. Age-dependent deterioration of nuclear pore complexes causes a loss of nuclear integrity in postmitotic cells. *Cell* **136**, 284–295 (2009).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank S. Michaelis for discussions; K. Russell, M. Elrick and J. Ravits for human tissue and/or human histological studies. We also thank the Hetzer Laboratory for the CMV–NLS–tdTomato–NES construct; B. Ganetzky and C. Staber for the *Drosophila* RanGAP antibody; P. Jin for the UAS-(G₄C₂)₃₀ fly stock; C. Svendsen for some control iPS cell lines; L. Petrucelli for the GP antibody; F. Hirth for the TBPH antibody. We thank the TriP at Harvard Medical School (NIH/NIGMS R01-GM084947) for providing transgenic RNAi fly stocks used in this study. Stocks obtained from the Bloomington *Drosophila* Stock Center (NIH P400D018537) were used in this study. This work was supported by grants from NIH (R01 NS085207 and NS091046 to J.D.R. and R.S., R01 NS082563 to T.E.L., R01 NS074324 and NS089616 to J.W.), Brain Science Institute, Robert Packard Center for ALS Research at Johns Hopkins, Muscular Dystrophy Association (J.D.R.), Alzheimer's Drug Discovery Foundation (J.D.R. and R.S.), Judith and Jean Pape Adams Charitable Foundation (J.W. and R.S.), Alzheimer's Disease Research Center – Johns Hopkins (R.S.), Maryland TEDCO (C.J.D. and J.W.), Target ALS Springboard Fellowship (C.J.D.), William and Ella Owens Foundation (R.S.), and ALS Association (T.E.L., R.S. and J.D.R.). K.Z. is a Milton Safenowitz fellow in the ALS Association. A.R.H. is a fellow on an NIH training grant (CA009110) and a recipient of a NIH K99 award (NS091486). J.C.G. and S.J.M. are recipients of a National Science Foundation Graduate Research Fellowship Award and J.C.G. is a recipient of the Thomas Shortman Training Fund Graduate Scholarship.

Author Contributions K.Z., C.J.D., R.S., T.E.L. and J.D.R. conceived the project. K.Z., C.J.D., A.R.H., J.W., R.S., T.E.L. and J.D.R. designed the experiments. K.Z. performed most studies related to *Drosophila*, with assistance from J.B.M., K.M.C. and S.G. C.J.D. performed studies employing iPSC neuronal cultures and human tissue with help from S.J.M., L.W.O. and J.C.G. A.R.H. performed the EMSA analysis. J.B.M. performed the fly NMJ and electrophysiological analyses; I.H., S.-L.C. and R.L.H. performed and/or interpreted the human iPS electrophysiological analyses. J.C.G., E.L.D., S.V., M.A.T. and P.S. provided technical support. A.R.H., K.Z. and C.J.D. developed the figures. K.Z., C.J.D., A.R.H., M.J.M., J.W., R.S., T.E.L. and J.D.R. interpreted data and prepared the manuscript. K.Z. and C.J.D. contributed equally to this work. T.E.L. and J.D.R. contributed equally to this work. All authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.D.R. (jrothstein@jhmi.edu) or T.E.L. (tlloyd4@jhmi.edu).

METHODS

For all *Drosophila* experiments, the experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment. Investigators were blinded for analysis of all iPSC neuron experiments investigating the N/C Ran ratio, N/C TDP-43 ratio, and FRAP live imaging.

Drosophila genetics. To identify genetic modifiers of G_4C_2 HRE, the candidate-based screen was performed as follows: if a candidate RBP⁸ is conserved between human and *Drosophila*, we obtained the RNAi lines against the *Drosophila* homologue(s) from the TRiP collection (Supplementary Table 1)³⁸. In addition, if the RBP consistently exhibited high affinity to G_4C_2 RNA, we also obtained published mutant alleles of their homologues. *RanGAP^{SD}* (*GOF*) refers to the *RanGAP^{SD}* 'segregation distortion' gain-of-function allele^{11–13}. We recombined *GMR-GAL4* and *UAS-(G₄C₂)₃₀* (ref. 10) and crossed the balanced line, *GMR-GAL4, UAS-(G₄C₂)₃₀/CyO, twiGFP*, to RNAi or mutant lines. We selected progeny that either co-expressed both the repeats and the RNAi (*GMR-GAL4, UAS-(G₄C₂)₃₀/+; UAS-RNAi/+*, where the *UAS-RNAi* can be on any chromosome), or expressed the repeats in a heterozygous mutant background (*GMR-GAL4, UAS-(G₄C₂)₃₀/+; mut/+*, where *mut* can be on any chromosome). We aged flies for 15 days and compared the morphology of their eyes with 15-day-old control flies expressing only the repeats. We used a modification index ranging from -4 to 4 to describe the relative severity of the morphological defects (Supplementary Table 1), where 0 is the repeat-expressing control. A positive number indicates enhancement of the phenotype, whereas a negative number indicates rescue. A number of 4 was given if the flies have no eyes, whereas a number of -4 was given if eyes appear indistinguishable from that of the wild-type control. If the flies fail to eclose, we indicate the phenotype as 'lethal'.

In our genetic interaction analyses, we used a previously described method³⁹ to quantify disruption in the external morphology of the eye, that is, 'degeneration score'. Briefly, points were added if there was complete loss of interommatidial bristles, necrotic patches, retinal collapse, loss of ommatidial structure, and/or depigmentation of the eye.

For the subcellular localization of GFP, *OK371-GAL4; UAS-(G₄C₂)₃₀/TM6b, Tb, tub::GAL80* was crossed to *UAS-NLS-NES(P12)/TM6b, Tb (III)* and non-*Tb* offspring were selected for analysis (*NES(P12)* is referred to as Δ NES). *OK371-GAL4/UAS-NLS-NES-GFP* flies were used as a negative control. We did not observe any GFP signals in the third instar salivary glands of *OK371-GAL4; UAS-(G₄C₂)₃₀/+* animals. All other fly stocks are from Bloomington *Drosophila* Stock Center, except for the *UAS-RanGAP* lines generated in this study.

To induce G_4C_2 RNA expression using *elavGS* (ref. 40), flies were raised at 29°C on regular food supplemented with $300\ \mu\text{M}$ RU486. Flies were transferred to freshly made food every 2–3 days.

Quantitative RT–PCR. For each genotype, mRNA was collected from 30 fly heads using the TRIzol reagent (Life Technologies) following the manufacturer's protocol. Reverse transcription was performed using SuperScript III First-Strand synthesis kit (Life Technologies) following the manufacturer's protocol. Quantitative PCR was performed using SYBR Green PCR system (Applied Biosystem) on a 7900 HT fast Real-Time PCR system (Applied Biosystem). The following primers were used. For *actin*: forward 5'-GCGCGGTTACTCTTTACCA-3', reverse 5'-ATGTACACGACGATTTACAG-3'. For G_4C_2 repeats: forward 5'-GGGATCTAGCCACCATG GAG-3', reverse 5'-TACCGTCGACTGCAGAGATTC-3'.

The primers for G_4C_2 repeats were designed to amplify a 3' region immediately after the repeats in the *UAS* construct.

Flight assay. The flight assay was performed as described⁴¹. Briefly, individual 15-day-old female flies were dropped into a graduated cylinder through a hole in its lid. The cylinder was graduated into 12 zones of 25 mm each (top: 0; bottom: 12). The landing height was noted as the zone number in which the fly landed.

Electrophysiological recording. For fly third instar larvae, neuromuscular (NMJ) recordings were performed from muscle 6 in segments A3 and A4 at room temperature in $1.5\ \text{mM}$ Ca^{2+} containing HL3 as described⁴².

For iPSC cells, whole-cell patch-clamp recordings were performed to assess the functionality of iPSC neurons. Neurons were perfused in HEPES-buffered extracellular solution ($143\ \text{mM}$ NaCl, $5\ \text{mM}$ KCl, $2\ \text{mM}$ CaCl_2 , $1\ \text{mM}$ MgCl_2 , $10\ \text{mM}$ HEPES, $10\ \text{mM}$ glucose, pH 7.2, 300 – $310\ \text{mOsm}$) in the presence of $1\ \mu\text{M}$ TTX and $20\ \mu\text{M}$ bicuculline. Whole-cell recording pipettes (4 – $8\ \text{M}\Omega$) were filled with a Cs-based internal solution ($115\ \text{mM}$ Cs-MeSO₄, $0.4\ \text{mM}$ EGTA, $5\ \text{mM}$ TEA-Cl, $2.8\ \text{mM}$ NaCl, $20\ \text{mM}$ HEPES, $3\ \text{mM}$ MgATP, $0.5\ \text{mM}$ Na₂GTP, pH 7.2, 290 – $300\ \text{mOsm}$) for voltage-clamp mEPSC recordings or with a K^+ -based internal solution ($2.7\ \text{mM}$ KCl, $120\ \text{mM}$ KMeSO₄, $9\ \text{mM}$ HEPES, $0.18\ \text{mM}$ EGTA, $4\ \text{mM}$ MgATP, $0.3\ \text{mM}$ Na₂GTP, $20\ \text{mM}$ phosphocreatine(Na), pH 7.3, $295\ \text{mOsm}$) for current-clamp experiments. Cells were held at $-70\ \text{mV}$ holding potential and recording was performed at room temperature. Signals were measured with MultiClamp 700B amplifier and digitized using a Digidata 1440A

analogue-to-digital board (Molecular Devices). Data acquisition was performed with pClamp 10.3 software and digitized at 5 or $20\ \text{kHz}$.

iPSC generation and differentiation to neurons. Patient fibroblasts were collected at Johns Hopkins Hospital with patient's consent (IRB protocol: NA_00021979) as described previously⁸. iPSC lines were created and initially characterized with an NIH-sponsored commercial agreement with iPierian (USA) using the 4 vector method. Sox2, Oct4, Klf4 and c-Myc encoding vectors were transduced into human fibroblasts using retrovirus delivery. Selected colonies were evaluated for expression of multiple pluripotent markers by quantitative PCR (qPCR) and/or immunocytochemistry. *In vitro* pluripotency was further determined by three germ layer differentiation via embryoid body formation. iPSCs were maintained in mTeSR1 (StemCell Technology) and passed once a week using dispase (StemCell Technology) following the manufacturer's instructions. Partially differentiated colonies were removed manually before differentiation analyses. The iPSCs were differentiated to neuroprogenitor cells, neurons and motor neurons via embryoid body (EB) formation by following the methods described previously (Supplementary Table 3)⁴³. At day 32 of differentiation, iPSC neurons were treated with $20\ \mu\text{M}$ Ara-C (Sigma) for 48 h to remove iPS glial progenitor cells and enrich for iPSC neurons. iPSC neuronal differentiation was confirmed by class-III Tubulin (Tuj1) immunostaining (Chemicon AB9354, 1:1,000), and cultures used for subsequent experiments were plated onto a confluent layer of mouse astrocytes, and analyses were performed at 55–69 DIV. Differentiation was assessed by immunofluorescence for the presence of MAP2-positive cells (SySy 188 004; 1:1,000) and neuronal morphology. Approximately 85–90% of cells were VGlut 1⁺ (SySy 135 303; 1:500), ~10% were VGAT⁺ (SySy 131 002; 1:500), ~40% of neuronal cultures were Islet-1⁺ (DSHB, 40.3A4; 1:50), ~90% were ChAT⁺ (Millipore, AB144P, 1:300). All lines were analysed at 50–70 DIV.

Molecular cloning. *RanGAP* full-length and/or truncated cDNAs were retrieved from cDNA clone LD16356⁴⁴ and subcloned into pUAS-attB vector⁴⁵ using BglII and NotI sites. An HA tag was added at the C terminus. The NLS–tdTomato–NES construct was provided by the Hetzer lab (Salk Institute) and was subcloned into PrecisionShuttle Lenti vector with C-terminal Myc-DDK tag (OriGen, catalogue number PS100064) using MluI and XhoI cloning sites.

Transgenic flies. Transgenic flies containing *RanGAP* cDNA constructs were generated by injecting the plasmid into *y w; PBac[yellow(+)-attP-3B]/VK00033* (chromosome III) embryos (BestGene, Inc.)⁴⁶.

Collection of human autopsied tissue. Human autopsied tissue used for these data are described in detail in Supplementary Table 2. The use of human tissue and associated decedents' demographic information was approved by the Johns Hopkins University Institutional Review Board and ethics committee (HIPAA Form 5 exemption, Application 11-02-10-01RD) and from the Ravitz Laboratory (UCSD) through the Target ALS Consortium.

Antisense oligonucleotide treatment for iPSC neuronal cultures. Modified 2'-methoxyethyl (MOE)/DNA antisense oligonucleotides were generated by Isis Pharmaceuticals. For antisense oligonucleotide treatment, antisense oligonucleotides were incubated in neural differentiation media (NDM) at $3\ \mu\text{M}$ then added to the iPSC neuronal cultures and replenished every 3 days for a total of 10 DIV with antisense oligonucleotides. Sequence for the antisense oligonucleotide-577061 targeted upstream of the G_4C_2 repeat is 'TACAGGCTGCGGTTGTTTCC' and the scrambled non-targeting control antisense oligonucleotide-141923 sequence is 'CCTTCCCTGAAGTTCTCTCC'.

RNA fluorescent *in situ* hybridization (FISH) and immunofluorescence. RNA-FISH of iPSC neurons was performed as previously described⁸. Briefly, 5' digoxigenin (DIG)-labelled locked nucleic acid FISH probes used were generated by Exiqon and targeted the GGGGCC repeat (CCCCG_{2.5}) (batch 611635) or a non-targeting scrambled probe (300514-04) as a control. Cell cultures were fixed in buffered 4% PFA, equilibrated in $1\times$ SSC with 40% formamide, incubated at 37°C for 10 min, then incubated with preheated probes (90 – 95°C) at 75°C for 35 min in 50% formamide and hybridization buffer containing $20\ \text{nM}$ of probe. Following hybridization, cells were washed $2\times$ with 50% formamide at 55°C for 15 min each and then washed with $2\times$ SSC 5 times for 5 min each.

Cells processed for RNA-FISH and protein immunofluorescence for RanGAP1 were treated once with Tris-glycine, and processed for standard protein immunofluorescence. Blocking buffer and immunofluorescence buffer consisted of 10% and 5% protease and heat-shocked BSA fraction V (Roche) in RNase free $1\times$ Tris buffered saline, respectively. To detect the DIG-labelled probe, an unconjugated mouse anti-DIG antibody (Jackson ImmunoResearch 200-002-156; 1:400) and RanGAP1 antibody (Santa Cruz sc-25630; 1:500) was used following by the appropriate secondary antibody (Jackson). Cells then underwent a series of 5 min washes with immunofluorescence buffer, Tris buffered saline, Tris-glycine, PBS with MgCl_2 and PBS, respectively. Cells were mounted onto slides with ProLong Antifade Gold mounting media with DAPI (Invitrogen).

Drosophila cell culture. S2 cells were cultured in Schneider's media supplemented with fetal bovine serum and antibiotics at 25 °C. The transfections were performed using Lipofectamine LTX (Life Technologies) following the manufacturer's instructions. G_4C_2 repeat mRNA was transcribed using the UAS/GAL4 system⁴⁷, driven by *Act-GAL4*. For immunofluorescent staining, cells were fixed 48 h after transfection. For actinomycin treatment to induce apoptosis, cells were treated with 0.7 μ M actinomycin D for 20 h.

Immunofluorescence, phalloidin staining and immunohistochemistry. For immunofluorescence staining in *Drosophila*, tissues or S2 cells were fixed in 3.7% formaldehyde for 30 min, followed by incubation in PBX solution (PBS with 0.4% Triton X-100) for 1 h. The tissues or cells were then incubated with primary antibodies and 10% normal goat serum (NGS) in PBX for 16 h at 4 °C. Primary antibodies were used at the following concentrations: mouse anti-Brp (DSHB), 1:100; mouse anti-GFP (Life technologies), 1:200; rat anti-HA (Roche), 1:200; rabbit anti-RanGAP (a gift from C. Staber, Stowers Institute), 1:500; mouse anti-Ran (BD Biosciences), 1:200; rabbit anti-TBPH (a gift from F. Hirth, King's College)⁴⁸, 1:200; and rabbit anti-antisense oligonucleotide (provided by F. Rigo, Isis Pharmaceuticals), 1:1,000. Next, samples were washed in PBX for 8 h at room temperature (RT) and then incubated with secondary antibodies conjugated to Alexa Fluor 546 or 488 (Life Technologies) in PBX+10% NGS at 4 °C for 16 h. The secondary antibodies were used at a dilution of 1:200. After that, samples are washed in PBX for 6 h and then stained with 1 μ M TO-PRO3 (Life Technologies) for 10 min at RT.

For phalloidin staining, fixed eyes were incubated in PBX with Alexa Fluor 488 Phalloidin (Life Technologies) at 1:20 for 16 h at 4 °C. The eyes were then washed in PBX for 1 h at RT before mounting.

For immunostaining of iPSC neurons, cells were grown on 12 mm coverglass on top of a confluent monolayer of mouse astrocytes fixed in 4% PFA, permeabilized in 0.3% Triton X-100/1 \times PBS and blocked in 10% normal donkey serum before incubation with primary antibody. For human autopsied tissue, paraffin motor cortex tissue (see Supplementary Table 2) was washed in xylene (3 \times 5 min), then a series of 100% ethanol (2 \times 5 min), 90% ethanol (1 \times 5 min), 70% ethanol (1 \times 5 min) and washed with water (2 \times 5 min). Antigen retrieval was performed using a steamer for 1 h in epitope retrieval solution (IHC world) then washed with water (3 \times 5 min). Slides were treated with 50:50 methanol: acetone solution for 10 min and then washed with 1 \times PBS (2 \times 5 min). Permeabilization was performed with 0.4% Triton X-100/1 \times PBS (8 min) and were then washed (1 \times PBS) and blocked overnight in 10% normal goat serum/1 \times PBS. Primary antibodies were added and incubated for 24 h at 4 °C (RanGAP1 Santa Cruz, sc-25630, 1:50; Nup205 Novus NBPI-91247, 1:50). For DAB staining, tissues were incubated with biotinylated goat-anti-rabbit secondary antibody (Jackson Immunoresearch) at 1:200 for 1 h at room temperature and then processed using the Vectastain Kit (Vector Labs) following the manufacturer's instruction. Cells and tissue were then washed 5 times for 5 min with 1 \times PBS and then 1 time with water and mounted using ProLong Antifade Gold with DAPI.

Microscopy and image analysis. For fly experiments, samples were mounted in Vectashield (Vector Laboratories) and analysed under a confocal microscope (model LSM510; Carl Zeiss) with its accompanying software using Plan Apochromat 63 \times , NA 1.4 objectives (Carl Zeiss) at RT. Images were captured by AxioCam HRc camera (Carl Zeiss) or Hamamatsu Flash 4.0 (Hamamatsu). Images are processed using ImageJ (National Institutes of Health). Deconvolution was performed using the Tikhonov–Miller method.

For iPSC experiments, Z-stack images taken on a Zeiss Axioimager with the Apotome tool or a Zeiss LSM700 (NIH Grant S10 OD016374) laser scanning confocal microscope, all images were taken at matched exposure times or laser settings and normalized within their respective experiment. All comparative images were processed using identical settings.

Nuclear/cytoplasmic ratios were quantified using Z-stacks of iPSC neuronal cultures on the Zeiss Axioimager with the Apotome tool. Full Z-stacks were taken at 0.5 μ m intervals and the individual planes were then projected into maximum intensity images removing any lower layers that contain the astrocyte monolayer. The nuclear region was determined using either DAPI or Lamin-B (Santa Cruz sc-6217; 1:300) and the cytoplasmic fraction was determined using MAP2 (SySy 188 004; 1:1000). Ran was visualized using a Ran antibody (BD Biosciences 610341; 1:100) or Ran-GFP (OriGene, RC204223L2). Images were quantified using ImageJ (NIH) and the mean pixel intensity per μ m² was determined to generate the nuclear/cytoplasmic ratios. All iPSC lines were imaged at 50–70 DIV. **FRAP analysis.** FRAP analysis was performed as previously described²⁷ with modifications. Ctrl or C9-ALS iPSC neurons were transduced with a lenti-CMV–NLS–tdTomato–NES construct provided by the Hetzer Laboratory (Salk Institute) at 51 DIV. Cells expressing the tdTomato reporter were then imaged on an LSM700 and processed with Zen software (Carl Zeiss). Three images were

taken of the tdTomato-expressing iPSC neurons at which point the nucleus was bleached for 30 iterations of 40–60% laser power and recovery was monitored every 3 s for 150 intervals. Recovery was normalized to the average of the pre-bleached signals. To account for global bleaching, all post-bleach signals were also normalized by a 'bleach factor' at each time point, which was determined by the per cent of signal lost post-bleach in an unbleached transfected cell.

Purification of recombinant RanGAP1 from *Escherichia coli*. RanGAP1 cDNA was provided by S. Blackshaw. The cDNA was cloned into a SspI-digested linear pET28a vector in frame with a 6 \times His-EGFP N-terminal fusion using Gibson assembly cloning strategies (NEB). A 50 ml LB starter culture of RanGAP1 with ampicillin/chloramphenicol was grown overnight at 37 °C. Then 25 ml of the overnight starter culture was added to 1 l of pre-warmed LB with ampicillin/chloramphenicol and incubated at 37 °C until OD₆₀₀ = 0.7–1 absorbance (abs). The temperature was then dropped down to 16 °C and protein expression was induced with 1 mM IPTG and induced overnight at 16 °C. The cell culture was then centrifuged at 4,000g for 20 min. The cell pellet was resuspended in 50 ml of resuspension buffer (20 mM HEPES, 200 mM NaCl, 10 mM imidazole, 1 mM TCEP, pH 7.4) containing EDTA-free protease inhibitor (Roche). Cells were lysed via French press while on ice, and were then centrifuged at 30,000g for 30 min. The lysed cell supernatant was collected and filtered through a 0.45 μ m membrane (Millex-HV Filter Unit) and then loaded onto an ÄKTA purifier 10 superloop at 4 °C. A 5 ml HisTrap HP Ni Sepharose (GE) column was pre-equilibrated with resuspension buffer before the supernatant was passed through. The protein was eluted off the column in elution buffer (20 mM HEPES, 200 mM NaCl, 500 mM imidazole, 1 mM TCEP, pH 7.4). Imidazole was removed by passing the protein solution through a HiTrap Desalting column (Sephadex G-25 Superfine, GE) pre-equilibrated in Desalting buffer (20 mM HEPES, 200 mM NaCl, 1 mM TCEP, pH 7.4). Removal of the His-GFP-TEV tag was facilitated by incubating the protein solution with 75 units of ProTEV Plus (5 U μ l⁻¹, Promega) per 2 ml of protein solution overnight while gently rocking at 4 °C. ProTEV Plus and the cleaved His-GFP-TEV tag were removed by reverse-Ni IMAC chromatography. The column flow through was collected and flash frozen in liquid nitrogen before storage at –80 °C. The flow through was then checked via SDS–PAGE gel and Coomassie stain to determine the purity of RanGAP1.

Electrophoretic mobility shift assays. A 24mer, 39mer or 60mer RNA (4 μ M) containing the sequences (G_4C_2), (C_4G_2), or (CUG) with a 5' Cy5 label (IDT) was denatured at 95 °C for 5 min and then annealed in the presence or absence of 100 mM KCl in 10 mM Tris-HCl pH 7.4 to induce the respective formation of RNA G-quadruplexes or hairpins. The RNA was diluted to 2 nM in binding buffer (HEPES pH 7.5 with 100 mM KCl, 5 mM MgCl₂, 50 μ M ZnCl₂, 1 mM TCEP, and 0.01% IGEPAL) and then incubated for 30 min at room temperature with varying concentrations of recombinant RanGAP1 (0, 1, 2, 10, 20, 50, 100 and 200 nM) in binding buffer. Samples were then loaded onto a 0.8% agarose gel in 1 \times TAE (pH 8.0) and electrophoresed for 45 min at 60 V. Bands were visualized using a Typhoon Image for Cy5 excitation and emissions. The image was analysed and quantified in ImageJ and then plotted in GraphPad prism. RanGAP1 binding was fit to a hyperbolic and linear regression, and based on the fit of the curve the $k_{1/2}$ calculated with B_{max} set to 1 for nonlinear regression.

Competition experiments were performed by incubating a final concentration of 2 nM of RNA with 100 nM of RanGAP1 for 30 min at room temperature as above. Then unlabelled competitor, antisense oligonucleotide control, antisense oligonucleotide, RNaseH-dependent antisense oligonucleotide, or yeast tRNA was added to the sample at increasing concentrations, and allowed to incubate at room temperature for an additional 30 min. Samples were then analysed as above.

The effects of the porphyrin TMPyP4 on RanGAP1 binding to the RNA was performed essentially as above. RNA (2 nM) was incubated with varying concentrations of TMPyP4 that was serially diluted tenfold starting from a 1 μ M final concentration in binding buffer. After 30 min incubation, 10 nM RanGAP1 was added and allowed to incubate for an additional 30 min; binding was analysed as described above.

Protein extraction, protein/RNA pull down and immunoblot. Tissues or cells were homogenized and/or lysed in RIPA buffer (50 mM Tris-HCl pH 7.4, 150 mM NaCl, 0.1% SDS, 0.5% sodium deoxycholate, and 1% Triton X-100) supplemented with protease inhibitor cocktail (Complete, Roche). For pull down, cells are lysed in lysis buffer (50 mM Tris pH 7.4, 150 mM NaCl, 1% NP-40, and 5 mM MgCl₂) for 30 min on ice. The lysate was then pre-cleared using avidin-agarose beads (Life Technologies) for 30 min before being incubated with biotinylated G_4C_2 -repeat RNA with 10 mM TCEP and RNase inhibitor RNaseOUT (Life Technologies)⁹. Protein/RNA mixture was then incubated with avidin-agarose beads overnight at 4 °C. The beads were subsequently precipitated by centrifuge at 1,500g for 3 min and washed three times in lysis buffer at 4 °C for a

total time of 1 h. The beads were then resuspended in 50 µl lysis buffer and subjected to immunoblot analysis.

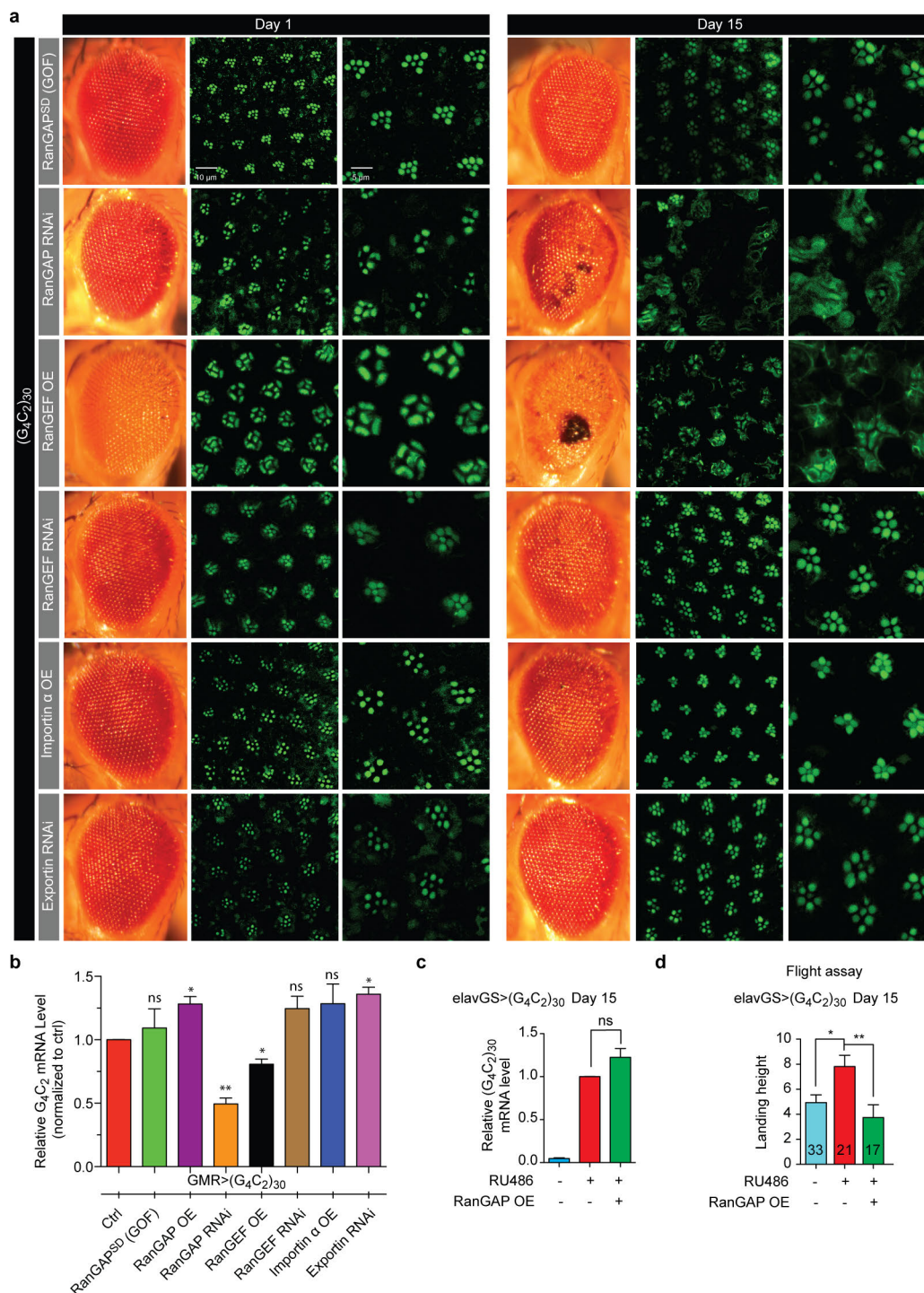
For immunoblot, the sample was mixed with Laemmli buffer and heated at 98 °C for 10 min. The protein samples were run on 4–15% SDS Mini-PROTEAN TGX Precast Gels (Bio-Rad) and transferred to nitrocellulose membrane. For dot blot, 2 µl of sample was blotted on nitrocellulose membrane and air-dried for 15 min. TBST with 5% milk was used for blocking. Primary antibodies were used as below: rat anti-HA (Roche), 1:1,000; chicken anti-GFP (abcam), 1:1,000; mouse anti-Actin (Millipore), 1:5,000; rabbit anti-GP (a gift from L. Petrucelli, Mayo Clinic), 1:1,000; and rabbit anti-GR (Proteintech), 1:1,000. The HRP-conjugated secondary antibody (Jackson ImmunoResearch) was used at 1:5,000 dilution.

Drug feeding assay. Melt cornmeal-molasses-yeast fly food was mixed with certain concentrations of antisense oligonucleotide 573674 'CCGGCCCCGG CCCCCGCCCC' (Isis Pharmaceuticals), TMPyP4 (a porphyrin derivative) (Sigma), or KPT-276 (Selleckchem) at high temperature and cooled to RT. PBS was used as the vehicle control for antisense oligonucleotide and DMSO was used as the vehicle control for TMPyP4 and KPT-276. Parent flies were crossed on food supplemented with drugs and the offspring were raised on the same food. Wandering third instars of the offspring were selected and subjected to GFP staining. Antisense oligonucleotides were detected using the anti-antisense oligonucleotide (13545) antibody (Isis Pharmaceuticals), which detects the MOE modification. Adult flies were aged on the drug-containing food for 15 days before analysing their eye morphology.

Statistics. No statistical methods were used to predetermine sample size. For quantification of outer eye morphological defects, ten flies were quantified. For quantifications of rhabdomere defects, 20 ommatidia from three or four flies were quantified for each genotype except RanGAP overexpression. For RanGAP overexpression, 24 ommatidia from four flies were quantified. For active zone quantifications, eight NMJs from four animals are quantified. For NMJ recording, the following numbers of animals were used for quantification: 18 for control, 10 for G₄C₂-expressing, 6 for RanGAP overexpression. For S2 cell quantifications, ten cells were quantified for each genotype. For image quantification of iPSC neurons, at least 31 neurons per line were quantified for all analyses and each cell line was differentiated and analysed at least two times at 55–70 DIV (as indicated) (see Supplementary Data Table 4). For qRT-PCR, six biological repeats, each containing three technical replicates in parallel, were used for quantification. For salivary gland quantifications, eight or nine salivary gland cells from three or four flies

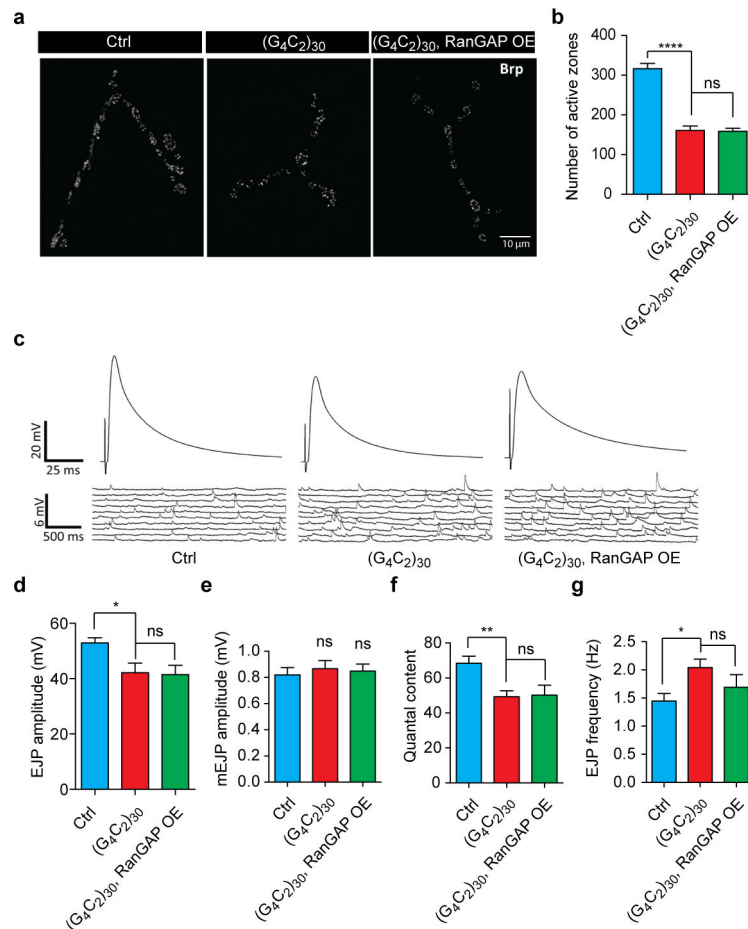
were quantified for each genotype. The numbers of flies used in the flight assay were labelled individually on the bar graph in Extended Data Fig. 1. Error bars are presented as s.e.m. For analyses of data sets with two variables, we employed a two-tailed Student's *t*-test. For analyses of data sets with three or more variables, we employed a one-way ANOVA assuming Gaussian distribution with a Tukey's post hoc test for multiple comparisons. To assess correlation between N/C ratios, a correlation analysis with Pearson's coefficient was applied. To obtain r^2 values, a nonlinear regression curve fit assuming one-phase association was performed. A *P* value of < 0.05 was considered statistically significant for all tests (GraphPad Prism, Ver. 6.0b).

38. Ni, J. Q. *et al.* A *Drosophila* resource of transgenic RNAi lines for neurogenetics. *Genetics* **182**, 1089–1100 (2009).
39. Ritson, G. P. *et al.* TDP-43 mediates degeneration in a novel *Drosophila* model of disease caused by mutations in VCP/p97. *J. Neurosci.* **30**, 7729–7739 (2010).
40. Osterwalder, T., Yoon, K. S., White, B. H. & Keshishian, H. A conditional tissue-specific transgene expression system using inducible GAL4. *Proc. Natl Acad. Sci. USA* **98**, 12596–12601 (2001).
41. Nelson, H. B. *et al.* Calmodulin point mutations affect *Drosophila* development and behavior. *Genetics* **147**, 1783–1798 (1997).
42. Machamer, J. B., Collins, S. E. & Lloyd, T. E. The ALS gene FUS regulates synaptic transmission at the *Drosophila* neuromuscular junction. *Hum. Mol. Genet.* **23**, 3810–3822 (2014).
43. Son, E. Y. *et al.* Conversion of mouse and human fibroblasts into functional spinal motor neurons. *Cell Stem Cell* **9**, 205–218 (2011).
44. Rubin, G. M. *et al.* A *Drosophila* complementary DNA resource. *Science* **287**, 2222–2224 (2000).
45. Bischof, J., Maeda, R. K., Hediger, M., Karch, F. & Basler, K. An optimized transgenesis system for *Drosophila* using germ-line-specific phiC31 integrases. *Proc. Natl Acad. Sci. USA* **104**, 3312–3317 (2007).
46. Venken, K. J., He, Y., Hoskins, R. A. & Bellen, H. J. P[acman]: a BAC transgenic platform for targeted insertion of large DNA fragments in *D. melanogaster*. *Science* **314**, 1747–1751 (2006).
47. Brand, A. H. & Perrimon, N. Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development* **118**, 401–415 (1993).
48. Diaper, D. C. *et al.* Loss and gain of *Drosophila* TDP-43 impair synaptic efficacy and motor control leading to age-related neurodegeneration by loss-of-function phenotypes. *Hum. Mol. Genet.* **22**, 1539–1557 (2013).
49. Devlin, A. C. *et al.* Human iPSC-derived motoneurons harbouring TARDBP or C9ORF72 ALS mutations are dysfunctional despite maintaining viability. *Nature Commun.* **6**, 5999 (2015).



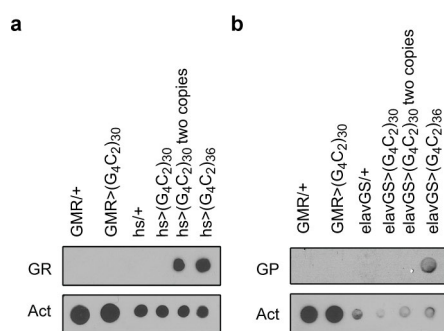
Extended Data Figure 1 | Genetic interaction between G_4C_2 repeats and components of the nucleocytoplasmic transport machinery. **a**, External eye morphology of 1-day-old (left column) and 15-day-old (left column) flies. Phalloidin staining of the retina of newly eclosed (middle column, magnified in right column) and 15-day-old (middle column, magnified in right column) flies is shown. Flies expressing 30 G_4C_2 repeats together with (from top row) RanGAP^{SD}(GOF), RanGAP RNAi, RanGEF overexpression, RanGEF RNAi, importin- α overexpression, or exportin RNAi. Genotypes (from top row): (1) *GMR-GAL4, UAS-(G_4C_2)₃₀/RanGAP^{SD}*; (2) *GMR-GAL4, UAS-(G_4C_2)₃₀/+*;

UAS-RanGAP RNAi/+; (3) *GMR-GAL4, UAS-(G_4C_2)₃₀/+*; *UAS-RanGEF/+*; (4) *GMR-GAL4, UAS-(G_4C_2)₃₀/UAS-RanGEF RNAi*; (5) *GMR-GAL4, UAS-(G_4C_2)₃₀/UAS-imp- α 2*; (6) *GMR-GAL4, UAS-(G_4C_2)₃₀/+*; *UAS-Exportin RNAi/+* (BL31353). **b**, **c**, Quantification of G_4C_2 mRNA levels by qRT-PCR. **d**, Flight assay. The top of the graduated cylinder is '0', and thus decreased landing height represents better flight ability. Genotypes (from left lane): (1) and (2) *UAS-(G_4C_2)₃₀/+*; *elavGS-GAL4/+*; (3) *UAS-(G_4C_2)₃₀/+*; *elavGS-GAL4/UAS-RanGAP*. Number of flies (*n*) tested indicated in column. **P* < 0.05; ***P* < 0.01.

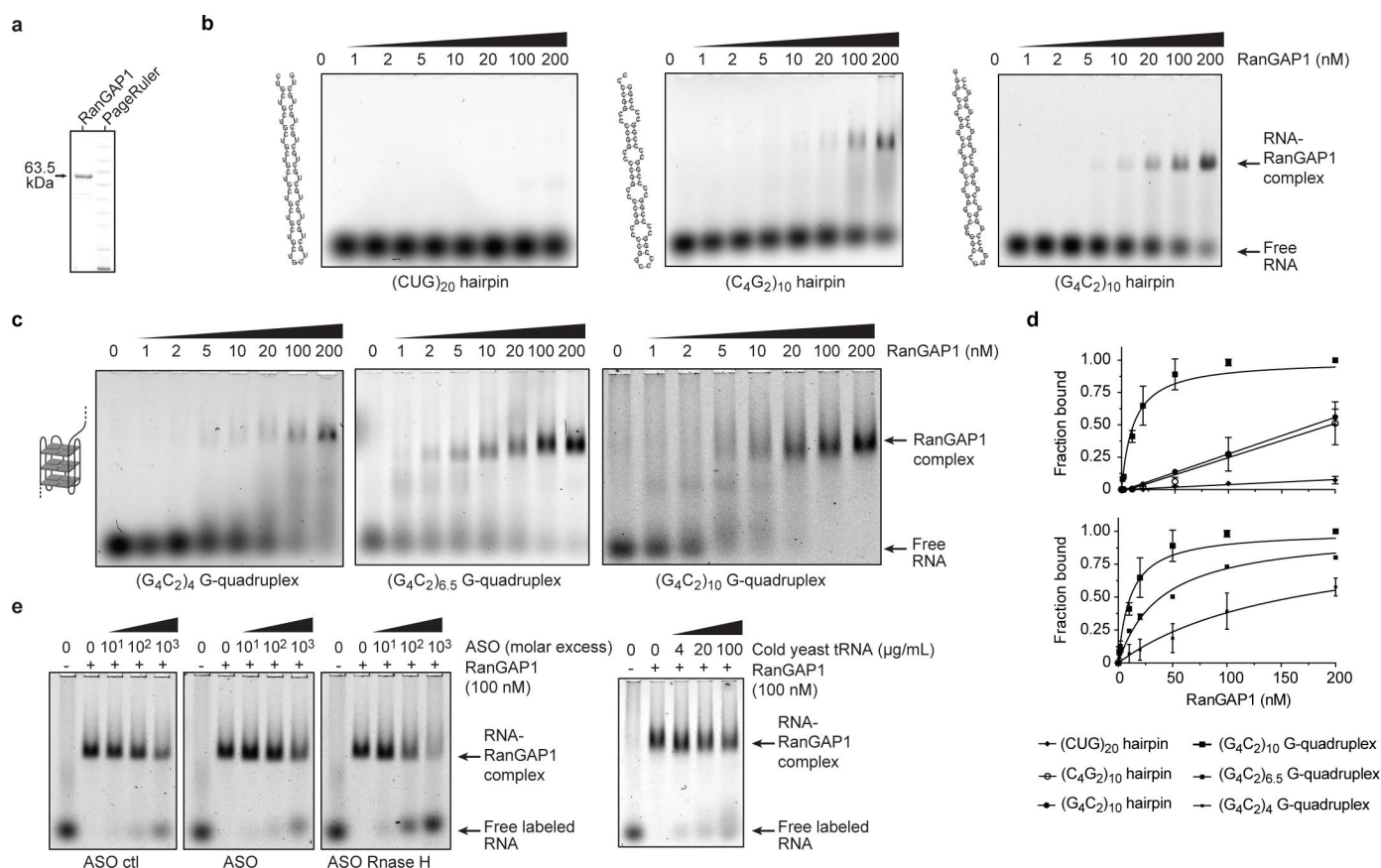


Extended Data Figure 2 | RanGAP does not rescue developmental defects caused by G₄C₂ repeats. **a**, Staining of the active zone component Bruchpilot (Brp) was used to identify active zones in the type Ib NMJ of muscle 4 in abdominal segments 3 and 4. **b**, Quantification of active zone number. **c**, Electrophysiological recording of NMJ in muscle 6/7 of abdominal segments

3 and 4. **d–g**, Evoked junctional potential (EJP) (**d**), miniature EJP (mEJP) amplitude (**e**), quantal content (**f**), and mEJP frequencies (**g**) are shown. Genotypes: (1) Ctrl, *OK371-GAL4/+*; (2) (G₄C₂)₃₀, *OK371-GAL4/+*; *UAS-(G₄C₂)₃₀/+*; (3) (G₄C₂)₃₀ RanGAP OE, *OK371-GAL4/+*; *UAS-(G₄C₂)₃₀/UAS-RanGAP*. **P* < 0.05; ***P* < 0.01; *****P* < 0.0001.



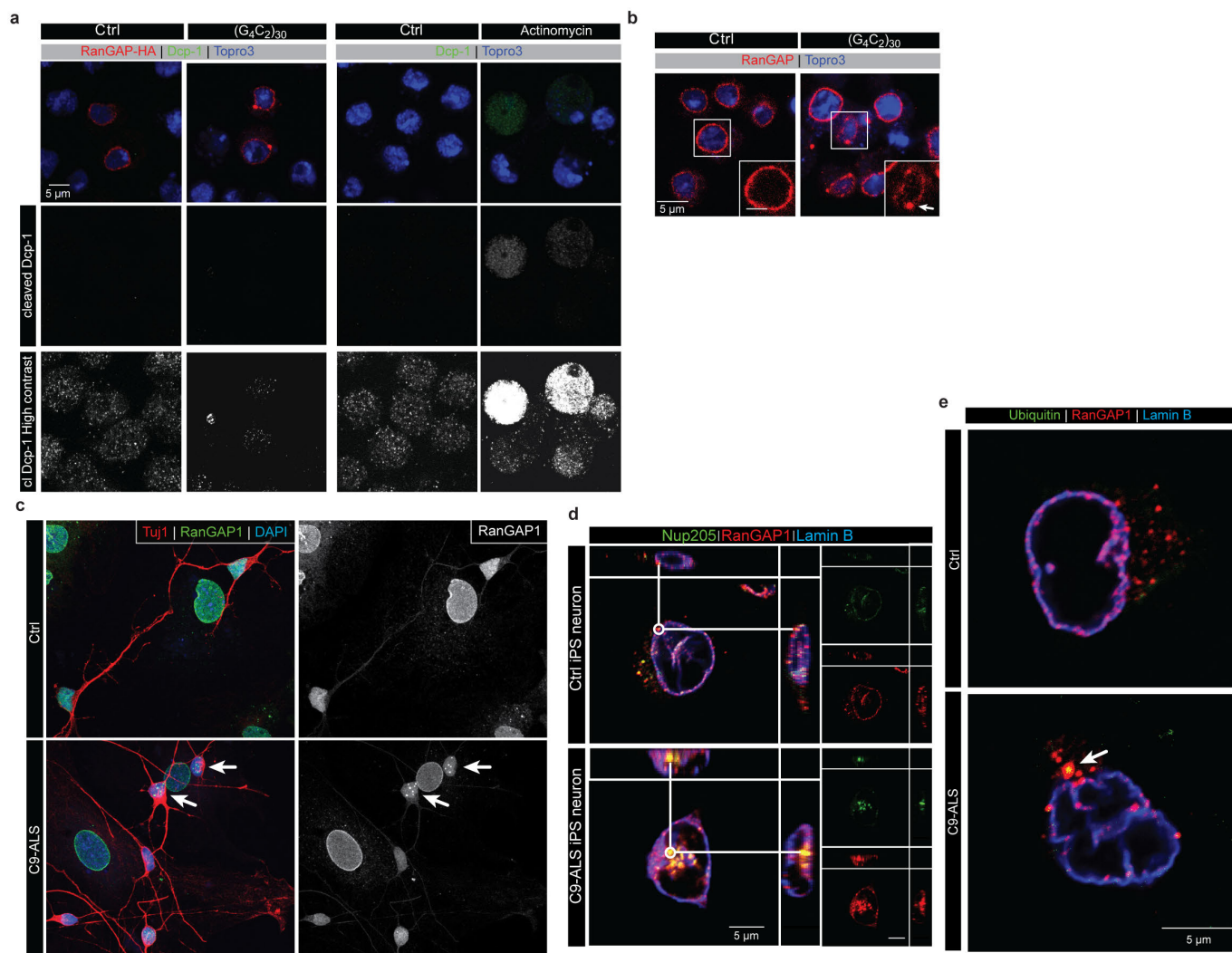
Extended Data Figure 3 | Dot blot of GR and GP dipeptide proteins. Dot blot of GR (**a**) and GP (**b**) compared with actin control. hs indicates heat-shock GAL4, and a heat shock was required to induce detectable polyGR as described⁷. A transgenic line *UAS-(G₄C₂)₃₆* previously shown to generate polyGR and polyGP DPRs under certain conditions was used as a positive control⁷.



Extended Data Figure 4 | RanGAP/RanGAP1 binds to G₄C₂ repeats.

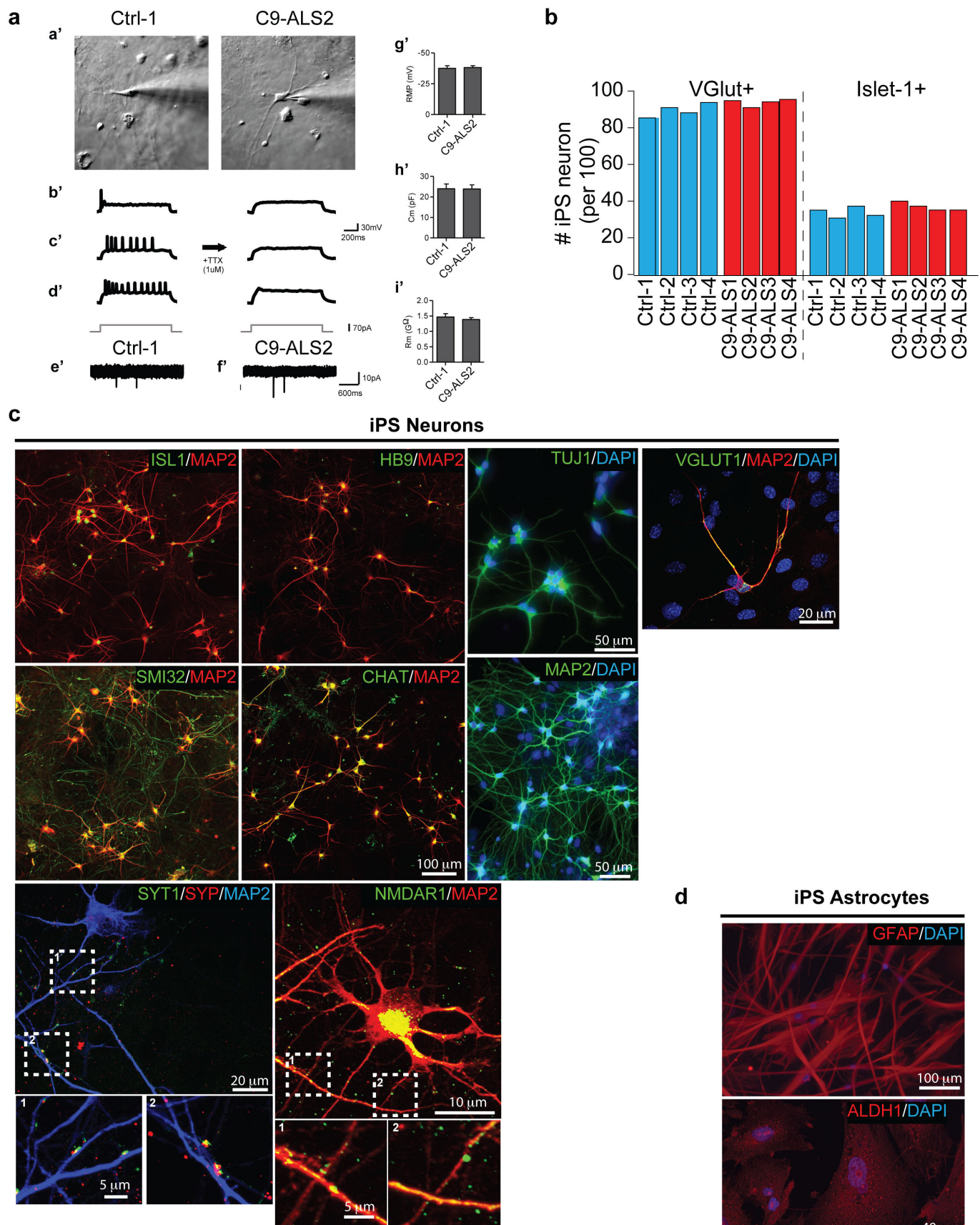
a, SDS-PAGE showing purified human RanGAP1. **b**, EMSA for RanGAP1 with (CUG)₂₀, (C₄G₂)₁₀, or (G₄C₂)₁₀ RNA hairpins. **c**, EMSA for RanGAP1 with increasing length of repeats that were annealed in the presence of K⁺ to promote RNA G-quadruplex formation. **d**, Plot of the fraction bound from the EMSAs performed with RanGAP1 and RNA repeats shown in **b** and **c**. Similar RNA nucleotide lengths but different binding preferences indicate that RanGAP1 has a structure- and sequence-dependent RNA binding mode (top panel). All data were fit using a hyperbolic and linear regression, then the RanGAP1 binding model determine based on the *r*² values for the best fit

(*n* = 2). The length-dependent binding of RanGAP1 fits best to a hyperbolic regression, which demonstrates specific binding to the (G₄C₂)_n G-quadruplex conformation, and the fraction bound increases with increasing nucleotide length (bottom panel). The fraction bound for the RNA hairpins fit best to a linear regression, which indicates nonspecific or less specific binding to RanGAP1. The *k*_{1/2} values for specific binding of RanGAP1 to the G-quadruplex RNA conformation are 162, 39 and 11 nM for (G₄C₂)₄, (G₄C₂)_{6.5} and (G₄C₂)₁₀, respectively. **e**, The RanGAP1–(G₄C₂)₁₀ RNA G-quadruplex complex is resistant to nonspecific RNA competitors and antisense oligonucleotides (*n* = 1).



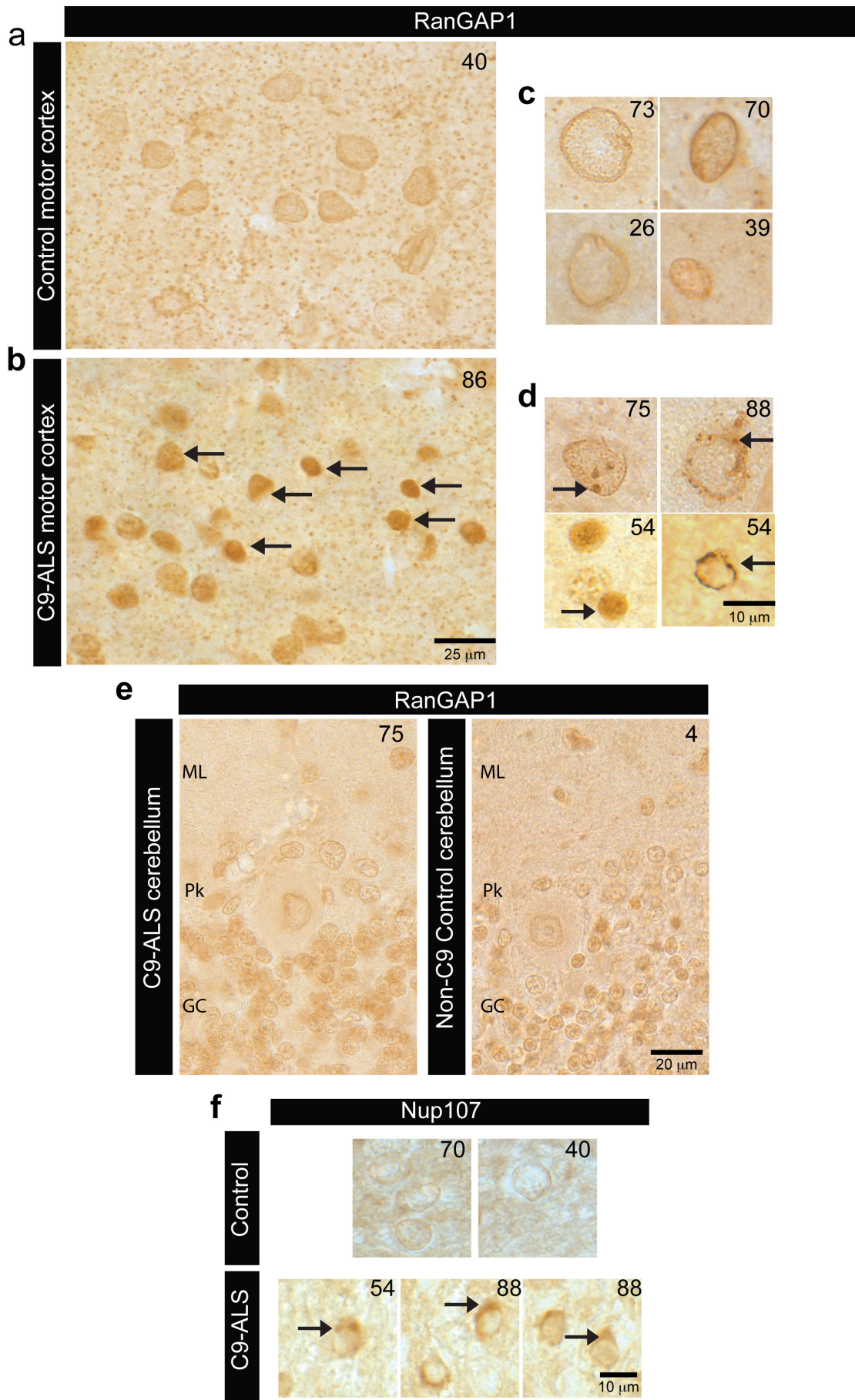
Extended Data Figure 5 | RanGAP/RanGAP1 is mislocalized in C9-ALS S2 and iPS cells. **a**, RanGAP mislocalization with (G₄C₂)₃₀ expression is not caused by apoptosis. S2 cells transfected with RanGAP-HA (first column) or RanGAP-HA and (G₄C₂)₃₀ (second column) were co-stained with HA (red), cleaved Dcp-1 (green) and TO-PRO3 (blue). As a control, S2 cells treated with DMSO (third column) or actinomycin (right column) are co-stained with cleaved Dcp-1 (green) and TO-PRO3 (blue). **b**, S2 cells transfected with G₄C₂ were co-stained with a Ran antibody (red) and TO-PRO3 (blue). **c**, Abnormal

aggregated RanGAP1 is variably observed in C9-ALS iPS neurons and is largely absent from control iPS neurons. Arrows indicate abnormal RanGAP1 staining. **d**, Single microscopic plane of aggregated RanGAP1 co-localized with Nup205 at the nuclear membrane (Lamin B) in C9-ALS iPS neurons. Single immunolabel view in right panels for Nup205 and RanGAP1, with *x-y* and *x-z* projections. **e**, Cytoplasmic RanGAP1 aggregates can co-localize with ubiquitin in C9-ALS iPS neurons.



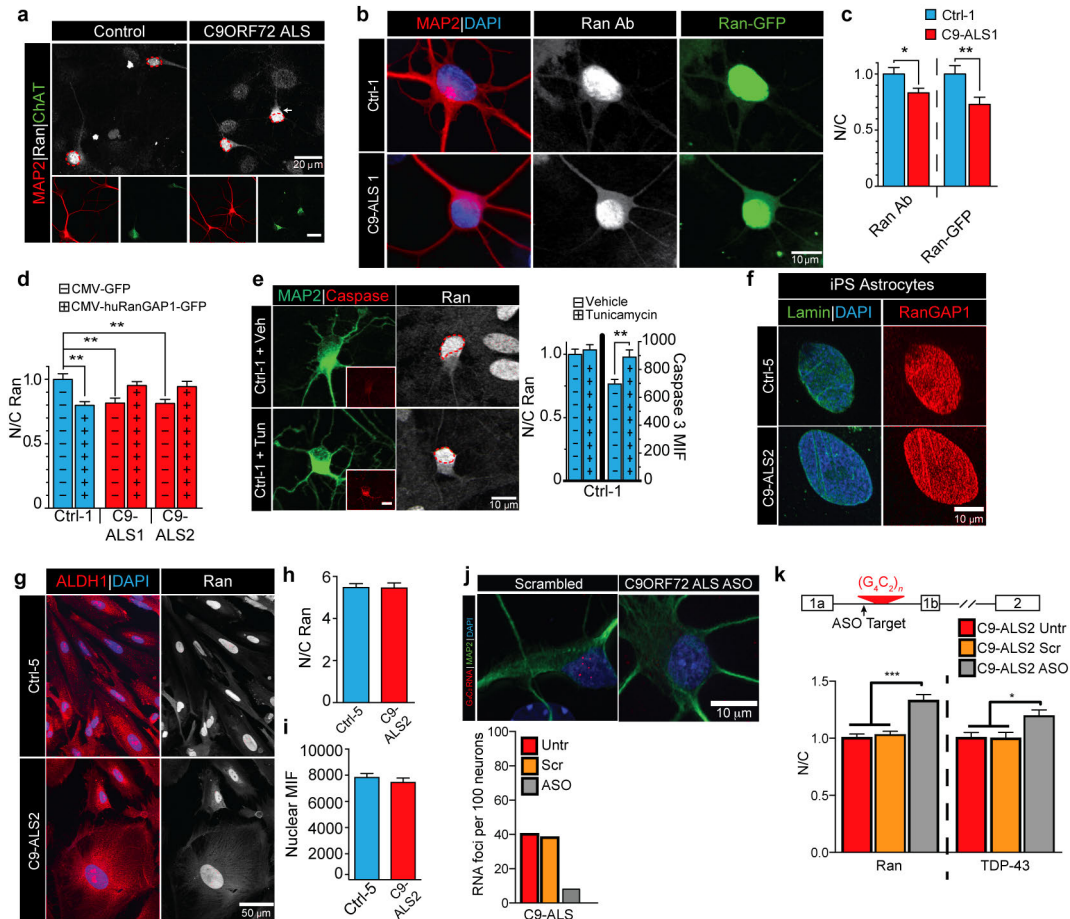
Extended Data Figure 6 | Electrophysiological and immunocytochemical characterization of iPSC neurons and astroglia. **a**, IR-DIC images of iPSC neurons from control (left panel) and *C9orf72* (right panel) patient cells (**a'**). Representative action potentials in response to somatic current injections (70 pA) in iPSC neurons (**b'–d'**). The majority of cells from both groups displayed either single, adaptive or repetitive responses, as demonstrated previously⁴⁹. These action potentials were blocked by TTX treatment. Normal (**e'**) and *C9orf72* (**f'**) patient cells displayed mEPSCs that were sensitive to NBQX treatment, suggesting functional synaptic input. Resting membrane

potential, membrane capacitance, and membrane resistance were comparable in both groups (**g'–i'**). **b**, Quantification of iPSC neuron markers showing glutamatergic and Islet-1⁺ iPSC neurons. **c**, iPSCs differentiated into neurons include phenotypic markers such as Islet-1, HB9, ChAT (choline acetyl transferase, motor neuron); Tuj1, MAP2, SMI32 (cytoskeletal), VGLUT1 (vesicular glutamate transporter 1), NMDAR1 (NMDA receptor), and synaptic markers SYT1 (synaptotagmin) and SYP (synaptophysin). **d**, Astroglia markers include ALDH1 (universal astroglial marker) and GFAP (reactive astroglia).



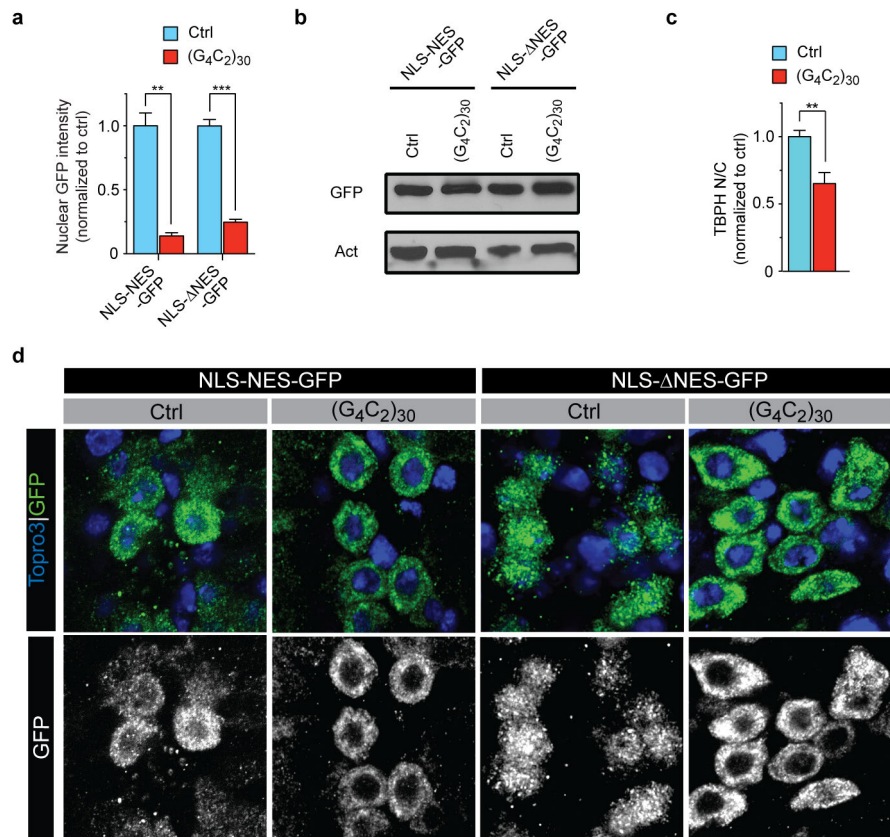
Extended Data Figure 7 | Additional human RanGAP1 and Nup107 pathology in C9-ALS brain. **a, b,** *C9orf72* motor cortex (**b**) reveals aberrant nuclear localization of RanGAP1, compared to a non C9 control tissue (**a**), including various nuclear aggregate pathologies seen at higher power in *C9orf72* ALS motor cortex (**d**) as compared to control (**c**). **e,** Aberrant RanGAP1 nuclear aggregates were not readily observed in

C9-ALS cerebellar cortex molecular layer (ML), Purkinje cells (PK) or granule cell (GL) layer when compared to non C9-ALS control cerebellum. Number in the upper right of each panel identifies autopsy specimen (Supplementary Table 2). **f,** Nup107 was also aggregated at the nuclear membrane in C9-ALS motor cortex cells when compared to non C9 control tissues.



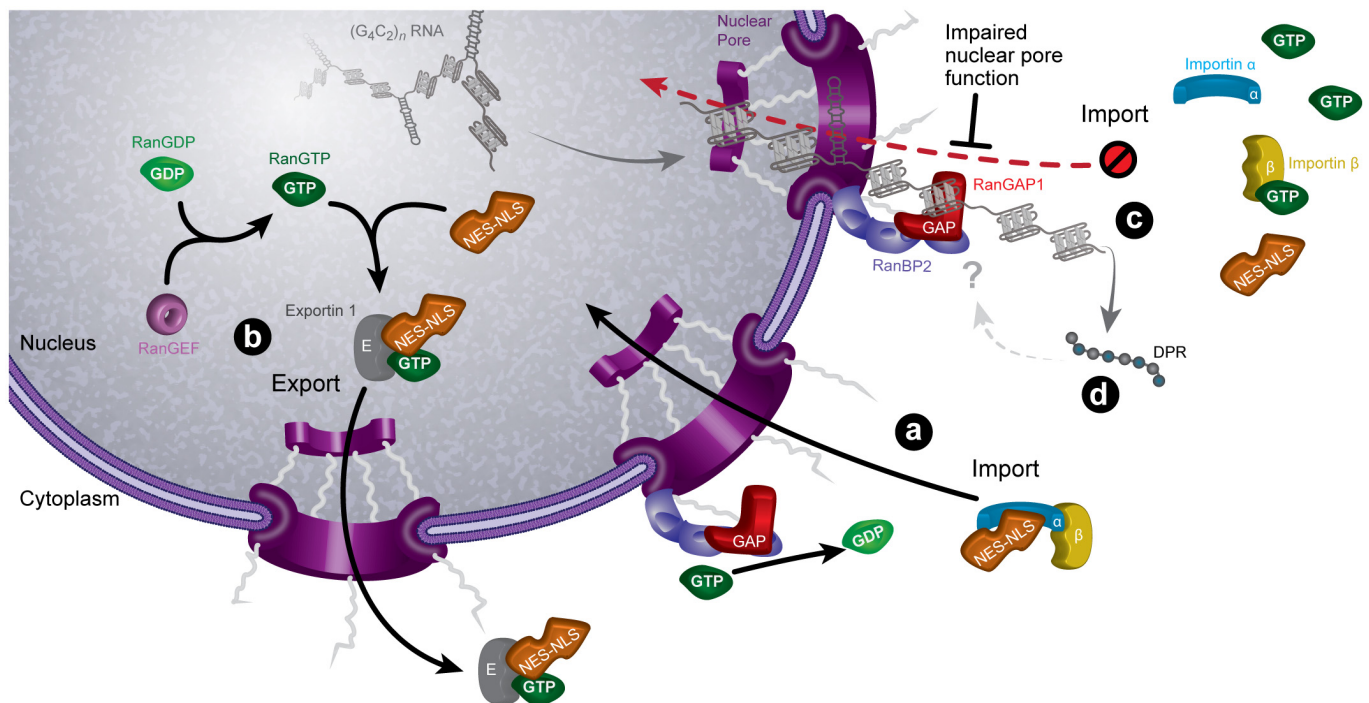
Extended Data Figure 8 | *C9orf72* HRE disrupts the cytoplasmic/nuclear Ran gradient. **a**, Representative images of disrupted N/C Ran gradient in C9-ALS ChAT⁺ iPSC neurons. **b**, **c**, Representative images and quantification of control (top row) or C9-ALS iPSC neurons (bottom row) expressing Ran-GFP that are co-stained with Ran and MAP2. Both Ran antibody and Ran-GFP indicate a reduced N/C Ran ratio. **d**, Overexpression of RanGAP1-GFP rescues the N/C Ran ratio in C9-ALS iPSC neurons. **e**, Control iPSC neurons treated with tunicamycin show enhanced level of activated Caspase 3 in the soma but no change in N/C Ran localization compared to controls with vehicle treatment. **f**, RanGAP1 is not aggregated in control and C9-ALS iPSC astroglia. **g**, Representative image of N/C Ran in C9-ALS astrocytes when

identified using the pan astroglial ALDH1 marker. **h**, N/C Ran is not altered in C9-ALS astroglia when comparing astrocytes of a similar size. **i**, Mean intensity fluorescence (MIF) of nuclear Ran does not differ in control or C9-ALS astroglia. **j**, Representative image of C9-ALS iPSC neuron with G₄C₂ RNA foci in approximately 40% of MAP2⁺ neurons at 50–70 DIV. Number of C9-ALS iPSC neurons with RNA foci is reduced with *C9orf72* RNA targeting antisense oligonucleotides compared to scrambled/non-targeting antisense oligonucleotides to <10% of iPSC neurons. **k**, Antisense oligonucleotides that reduce G₄C₂ RNA foci also enhance N/C Ran and N/C TDP-43 ratios. **P* < 0.05; ***P* < 0.01; *****P* < 0.0001.



Extended Data Figure 9 | C9orf72 HRE causes nucleocytoplasmic transport defects. **a**, Quantification of the nuclear GFP intensity in Fig. 4a. **b**, Immunoblot of the GFP levels in Fig. 4a. **c**, Quantification of the TBPH N/C ratio in Fig. 4a. **d**, Wild-type control and (G₄C₂)₃₀-expressing motor neurons expressing NLS-NES-GFP (left two columns) or NLS-ΔNES-GFP (right two columns) co-stained with a GFP antibody (green) and TO-PRO3

(blue) (top row). The GFP signal is shown separately in the bottom row. Genotypes (from left): (1) OK371-GAL4/UAS-NLS-NES-GFP (II); (2) OK371-GAL4/UAS-NLS-NES-GFP; UAS-(G₄C₂)₃₀/+; (3) OK371-GAL4/+; UAS-NLS-NES(P12)-GFP/+; (4) OK371-GAL4/+; UAS-NLS-NES(P12)-GFP/UAS-(G₄C₂)₃₀.



Extended Data Figure 10 | Model of *C9orf72* mutation induced nucleocytoplasmic transport disruption. **a**, In normal cases, RanGAP1 is tethered onto the NPC via RanBP2, where it activates Ran•GTP hydrolysis to produce Ran•GDP. Ran•GDP dissociates from and activates the Importin-αβ complex to import NLS–NES-containing protein cargoes such as TDP-43. **b**, In the nucleus, RanGEF converts Ran•GDP to Ran•GTP that is required for the dissociation of the NLS–Importin-αβ complex and the export of NES protein cargoes. **c**, In C9-ALS, G4C2 HRE binds and sequesters RanGAP1, leading to an increase in cytoplasmic Ran•GTP. High cytoplasmic Ran•GTP prevents the formation of the NLS–Importin-αβ complex, thereby disrupting the N/C Ran gradient and impairing nuclear import of NLS-containing proteins. **d**, Dipeptide repeat proteins translated from the G4C2 RNA can be toxic when expressed at high levels but it is unclear whether they contribute

to nucleocytoplasmic trafficking deficits in *Drosophila* since they are not detected at the time of degeneration. The *C9orf72* HRE sense strand appears to be contributing to nucleocytoplasmic trafficking deficits in human iPSC neurons and fly model systems, as small molecules and antisense oligonucleotides targeting the sense RNA substantially suppress the nuclear import phenotypes and neurodegeneration as a result of the G4C2 repeat RNA expression. Overall, the data are most consistent with an RNA-mediated mechanism with evidence that includes: (1) RanGAP1 was identified as 1 of 19 sequence-specific interactors of G4C2 RNA; (2) RanGAP is a strong genetic modifier of G4C2 RNA-mediated degeneration in *Drosophila* under conditions in which polyGR and polyGP are not detected; (3) RanGAP directly and potentially interacts with HRE RNA; and (4) G4C2 RNA foci can co-localize with RanGAP1.

Architecture of the synaptotagmin–SNARE machinery for neuronal exocytosis

Qiangjun Zhou^{1,2}, Ying Lai^{1,2*}, Taulant Bacaj^{1*}, Minglei Zhao^{1,2*}, Artem Y. Lyubimov^{1,2}, Monarin Uervirojnangkoorn^{1,2}, Oliver B. Zeldin^{1,2}, Aaron S. Brewster³, Nicholas K. Sauter³, Aina E. Cohen⁴, S. Michael Soltis⁴, Roberto Alonso-Mori⁴, Matthieu Chollet⁴, Henrik T. Lemke⁴, Richard A. Pfuetzner^{1,2}, Ucheor B. Choi^{1,2}, William I. Weis⁵, Jiajie Diao^{1,2}, Thomas C. Südhof¹ & Axel T. Brunger^{1,2}

Synaptotagmin-1 and neuronal SNARE proteins have central roles in evoked synchronous neurotransmitter release; however, it is unknown how they cooperate to trigger synaptic vesicle fusion. Here we report atomic-resolution crystal structures of Ca^{2+} - and Mg^{2+} -bound complexes between synaptotagmin-1 and the neuronal SNARE complex, one of which was determined with diffraction data from an X-ray free-electron laser, leading to an atomic-resolution structure with accurate rotamer assignments for many side chains. The structures reveal several interfaces, including a large, specific, Ca^{2+} -independent and conserved interface. Tests of this interface by mutagenesis suggest that it is essential for Ca^{2+} -triggered neurotransmitter release in mouse hippocampal neuronal synapses and for Ca^{2+} -triggered vesicle fusion in a reconstituted system. We propose that this interface forms before Ca^{2+} triggering, moves en bloc as Ca^{2+} influx promotes the interactions between synaptotagmin-1 and the plasma membrane, and consequently remodels the membrane to promote fusion, possibly in conjunction with other interfaces.

Membrane fusion is essential for many physiological processes in eukaryotic cells, including protein and membrane trafficking, hormone secretion and neurotransmitter release^{1,2}. Evolutionarily conserved SNARE (soluble *N*-ethylmaleimide sensitive factor attachment protein receptor) proteins have a key role in these processes. Specific combinations of SNARE proteins are located on opposite membranes. Upon zippering into a highly stable four-helix bundle—the SNARE complex—they provide the energy for membrane fusion^{3,4}. However, other factors are essential for regulation of membrane fusion. In particular, several proteins are required for neurotransmitter release in addition to neuronal SNAREs⁵, but it is unknown, at the atomic level, how these factors cooperate with SNAREs to promote synaptic transmission. One key factor is the Ca^{2+} sensor synaptotagmin, which consists of a short N-terminal luminal segment, a single transmembrane α -helix, an unstructured linker, and two Ca^{2+} -binding C2 domains, termed C2A and C2B, respectively (or C2AB together)⁶. There are 16 isoforms of mammalian synaptotagmins that are localized to synaptic and secretory vesicles or the plasma membrane. Among these isoforms, synaptotagmin-1 (Syt1) is a Ca^{2+} sensor for evoked synchronous neurotransmitter release⁷. Synaptotagmin-2 and synaptotagmin-9 are also involved in evoked synchronous neurotransmitter release for different subsets of neurons⁸. In contrast, synaptotagmin-7 plays a part in ‘slower’ asynchronous release^{9,10}; moreover, these and other synaptotagmins act in other types of exocytosis⁵. In addition to its role in evoked synchronous release, Syt1 also clamps the frequency of miniature spontaneous events^{11–13}.

Syt1 binds in a Ca^{2+} -dependent manner to anionic membranes; during binding, anionic phospholipids and synaptotagmin C2

domains together coordinate calcium ions^{14–17}. The membrane–synaptotagmin interaction has functional significance since the Ca^{2+} affinity of Syt1 for binding to anionic membranes and the Ca^{2+} sensitivity of neurotransmitter release are tightly correlated^{16,18}. The Syt1 C2AB fragment can induce vesicle clustering¹⁹ and preferentially binds to curved membranes^{20,21}. Moreover, C2 domains may penetrate the membrane upon Ca^{2+} binding^{22,23}.

Syt1 also interacts with the neuronal SNARE complex based on immunoprecipitation and pull-down experiments^{24–27}, single molecule fluorescence resonance energy transfer (smFRET)²⁸, and nuclear magnetic resonance^{29,30} experiments. A gain-of-function mutation in the Ca^{2+} -binding region of the C2A domain suggested that the Syt1–SNARE interaction may be functionally important²⁶, but the molecular basis and the significance of the interaction between Syt1 and the SNARE complex remain unknown.

Several crystal structures of Syt1 C2A and C2B domains, and C2AB fragments, are available^{31–33}, as well as the structure of the neuronal SNARE complex³; however, the atomic-resolution structure of the complex between Syt1 and the neuronal SNARE complex (referred to as Syt1–SNARE complex) has been elusive. Single molecule methods allowed the study of the Syt1–SNARE complex under dilute conditions with spatially isolated neuronal SNARE complexes reconstituted in a supported bilayer²⁸. The observed smFRET histograms²⁸ suggested several possible interfaces between Syt1 and the SNARE complex. Other dynamic or approximate models of the C2AB–SNARE complex were obtained by nuclear magnetic resonance (NMR)^{29,30}, but cannot be readily compared with the previous smFRET studies²⁸ or the results presented here because of differences

¹Department of Molecular and Cellular Physiology, Howard Hughes Medical Institute, Stanford University, Stanford, California 94305, USA. ²Departments of Neurology and Neurological Sciences, Photon Science, and Structural Biology, Stanford University, Stanford, California 94305, USA. ³Physical Biosciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA. ⁴SLAC National Accelerator Laboratory, Stanford, California 94305, USA. ⁵Departments of Structural Biology, Molecular and Cellular Physiology, and Photon Science, Stanford University, Stanford, California 94305, USA.

*These authors contributed equally to this work.

in conditions, particular covalent attachment of lanthanide labels³⁰, and lack of atomic resolution.

Here we report atomic-resolution crystal structures of a Syt1–SNARE complex in two different crystal forms and in the presence of either Ca^{2+} or Mg^{2+} . We found several interfaces, including a large structurally and evolutionarily conserved interface that is Ca^{2+} independent. Structure-based mutations of this interface disrupt evoked neurotransmitter release in primary neurons and Ca^{2+} -triggered fusion in a reconstituted system.

Structure of the Syt1–SNARE complex

We designed and tested several chimaeric constructs involving the Syt1 C2AB fragment (amino acids 141–421) and the neuronal SNARE complex (Extended Data Fig. 1a, b and Methods). We crystallized the Ca^{2+} - and Mg^{2+} -bound Syt1–SNARE complexes (Extended Data Fig. 1d, e), and determined their structures (Fig. 1 and Extended Data Table 1). The crystallization conditions were at near-physiological pH and ionic strength (Methods and Extended Data Fig. 1e). We observed two crystal forms for the Ca^{2+} -bound Syt1–SNARE complex, referred to as ‘short unit cell’ and ‘long unit cell’ crystal forms hereafter (Extended Data Figs 2 and 3).

The X-ray free-electron laser (XFEL) of the Linac Coherent Light Source (LCLS) at SLAC National Accelerator Laboratory yielded substantially higher-quality diffraction data than the Advanced Photon Source (APS) NE-CAT microfocus synchrotron beamline at Argonne National Laboratory from similar crystals of the long unit cell crystal form (Extended Data Fig. 2a, b). The electron density maps obtained from the XFEL diffraction data were notably superior to those of the synchrotron data sets (Methods). In particular, the electron density maps calculated from the XFEL data set were of sufficient quality to obtain accurate rotamers for most side chains, including those at the interfaces between molecules (Extended Data Fig. 2d–f). This is one of the first new crystal structures determined using XFEL diffraction data. Moreover, in contrast to the tens of thousands to millions of crystals typically used in XFEL-based crystallography experiments³⁴,

we obtained a reasonably complete data set from a few hundred images captured from 72 of the 148 crystals exposed to the LCLS XFEL beam (Methods).

Three distinct Syt1–SNARE interfaces

The crystal structures of the Syt1–SNARE complex reveal three interfaces between the SNARE complex and the Syt1 C2A and C2B domains (referred to as ‘primary’, ‘secondary’ and ‘tertiary’), as well as an interface between Syt1 C2 domains (referred to as C2A–C2B interface; Fig. 1a, b and Supplementary Videos 1 and 2). There are two essentially identical instances of the primary interface formed between Syt1 C2B domains and the SNARE complex in the long unit cell crystal form (Fig. 1c, d). The secondary interface involves another Syt1 C2B domain and the SNARE complex, while the tertiary interface involves a Syt1 C2A domain and the SNARE complex (Fig. 1b). These two C2 domains also form the C2A–C2B interface. All three interfaces between Syt1 and the SNARE complex fall within the range of smFRET efficiency histograms obtained previously²⁸ (Extended Data Fig. 4). These interfaces may suggest how multiple Syt1 and SNARE complexes simultaneously interact in the neuron (see below).

The largest, primary interface between the Syt1 C2B domain and the SNARE complex is very similar in both complexes (Fig. 1b) and in both crystal forms (Extended Data Fig. 5a), suggesting that it is not affected by crystal packing. The primary interface is also very similar in the Ca^{2+} -bound as well as in the Mg^{2+} -bound crystal structures (Extended Data Fig. 5d), implying a Ca^{2+} -independent interface. The residues involved in the primary interface have relatively low temperature factors, among the lowest in these structures (Extended Data Fig. 5e), and the electron densities of the side chains that form this interface are well defined (Fig. 2c, d), suggesting a specific interaction. In a recent study, lanthanide labels were covalently attached to SNAP-25 residues 41 and 166 for pseudo-contact chemical shift NMR measurements between Syt1 and the SNARE complex in the presence of 125 mM thiocyanate³⁰; these covalent labels would probably disrupt the primary interface (Fig. 2c), so it is not possible to compare our

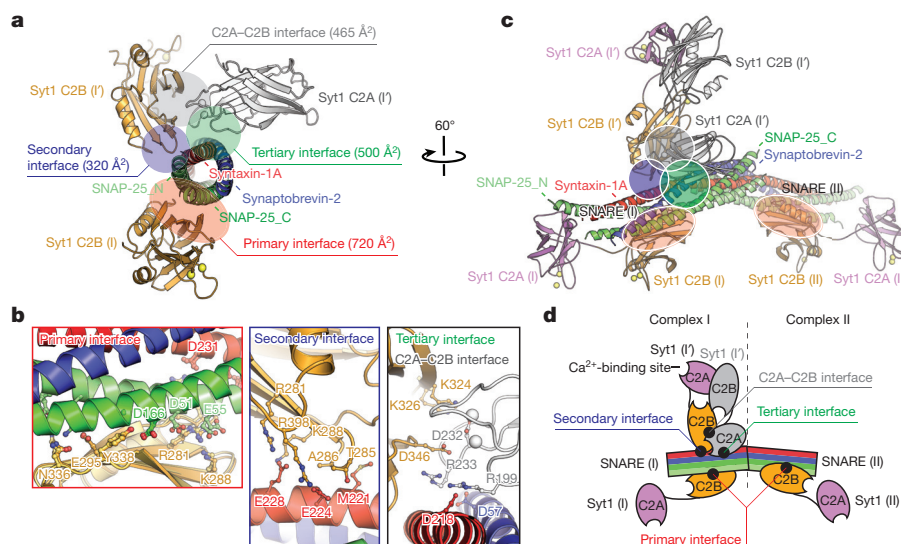


Figure 1 | Crystal structure of the Syt1–SNARE complex. **a**, Structure of the Ca^{2+} -bound Syt1–SNARE complex (only showing complex I) in the long unit cell crystal form (Extended Data Table 1 and Extended Data Fig. 2). Two Syt1 C2B domains (designated as I and I') and one Syt1 C2A domain (related by crystallographic symmetry and coloured in grey) form a total of three interfaces (primary, secondary and tertiary) with the neuronal SNARE complex (syntaxin-1A and the two SNARE domains of SNAP-25A (SNAP-25_N and SNAP-25_C)). A fourth interface (C2A–C2B interface) is located between Syt1 C2B (I') and Syt1 C2A (I'). Interface areas are provided in parentheses. **b**, Close-up views of the four interfaces with labels for side chains of interacting residues. The left panel shows a superposition of both

primary interfaces that occur in the long unit cell crystal form (root-mean-square difference (r.m.s.d.) = 0.34 Å, including C α atoms of the SNARE complex and the Syt1 C2B domain forming the interface). The middle panel shows the secondary interface. The right panel shows both the tertiary interface and the C2A–C2B interface. **c**, Rotated view of panel **a**, but showing the entire asymmetric unit and the symmetry-related Syt1 C2AB fragment. Three Syt1 C2AB fragments (designated as I, I' and II) bind to two SNARE complexes in the asymmetric unit. SNARE (II) only interacts with the C2B domain of one Syt1 C2AB fragment, Syt1 (II), via the same primary interface as observed in complex I. **d**, A schema corresponding to the structure shown in panel **c**.

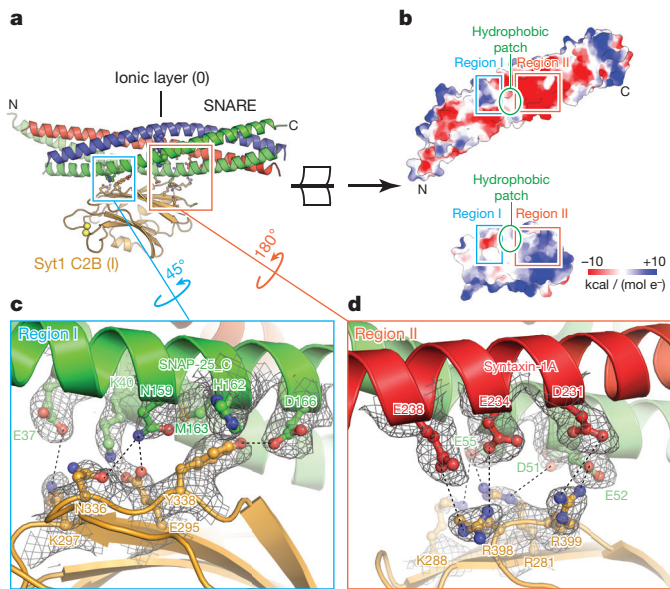


Figure 2 | Primary interface between the Syt1 C2B domain and the neuronal SNARE complex. **a**, Overview of the primary interface (complex I in the long unit cell crystal form) along with interacting residues (stick and ball representation). **b**, Open-book view of the electrostatic potential map of the primary interface. The two polar regions I and II are connected by a hydrophobic patch (SNAP-25 I44, L47 and V48 and Syt1 V292, L294 and A402). **c**, **d**, Close-up views of regions I and II. Interacting residues are labelled, along with dashed lines that indicate hydrogen bonds or salt bridges. $2mF_o - DF_c$ electron density maps of the interacting residues are superimposed (grey mesh; contour level = 1.5σ).

crystal structures with this NMR study. The interacting residues of the primary interface are conserved across different species for synaptotagmins (Syt1, Syt2, Syt9; ref. 8), SNAP-25 and syntaxin-1A homologues involved in fast synchronous release (Extended Data Fig. 6). In contrast, the interacting residues show variation among other synaptotagmin, SNAP-25 and syntaxin homologues that are not known to be involved in fast synchronous release.

The C2B domain forming the secondary interface is slightly rotated with respect to the SNARE complex between the two crystal forms and between the Ca^{2+} - and Mg^{2+} -bound crystal structures (Extended Data Fig. 5b, d), although the interactions at the secondary interface itself are similar. The C2A domain that forms the tertiary interface is in the same orientation in both crystal forms and in the Ca^{2+} - and Mg^{2+} -bound crystal structures (Extended Data Fig. 5c, d), and the interacting side chains are in similar positions. The interactions that form the tertiary interface are primarily ionic, involving residues Syt1 R199 and R233, syntaxin-1A D218, and synaptobrevin-2 D57 (Fig. 1b) with well-resolved electron density.

The conserved primary interface

We divide the primary interface into two regions (Fig. 2) dominated by polar interactions: region I comprises SNAP-25 residues E37, K40, N159, M163 and D166, and Syt1 residues E295, K297, N336 and Y338; region II comprises SNAP-25 residues D51, E52 and E55, syntaxin-1A residues D231, E234 and E238, and Syt1 residues R281, K288, R398 and R399. The SNARE complex is positively charged and Syt1 is negatively charged in region I, whereas the opposite is observed for region II. In contrast, other synaptotagmin isoforms not involved in fast synchronous release exhibit sequence variations and differences in the electrostatic potential maps in both regions (Extended Data Fig. 6a, b), suggesting that the primary interface may be an interaction that is specific for fast synchronous release.

A subset of the interacting residues of region II were suggested to be functionally important in previous studies: Syt1 R398/R399 (refs 35,

36) and SNAP-25 D51/E52/E55 (ref. 27). However, the role of these residues was unclear in the absence of structural information, which now reveals that they are part of a larger interface between Syt1 and the SNARE complex. Moreover, interactions in region I (Fig. 2c) have not been implicated in any previous studies.

The primary interface is critical

We designed mutations of the critical interacting residues of the primary interface based on the crystal structure and verified that all mutants result in properly folded Syt1 and SNARE complex (Methods and Extended Data Fig. 7). To test the interactions in neurons, we performed co-immunoprecipitation of syntaxin-1A in cultured Syt1 conditional knockout neurons infected with viruses expressing wild-type Syt1 or Syt1 containing the designed C2B mutants (Fig. 3a and Methods). The region II Syt1 mutant (R398Q/R399Q) has been reported previously^{35,36}, and is used here for comparison. Syntaxin-1A was immunoprecipitated from lysates of these neurons, and the presence of co-immunoprecipitated proteins was assayed with monoclonal antibodies against Syt1 and synaptobrevin-2. Mutation of either regions I and II of the primary interface reduced Syt1 binding to the SNARE complex by ~50%, and simultaneous mutation of both regions (referred to as 'Syt1 quintuple') reduced binding to ~33% as compared to a Syt1 wild-type construct introduced in the same manner (Fig. 3b, c). These results suggest that both regions are required for efficient Syt1–SNARE binding.

For further investigation of the function of the primary interface between the neuronal SNARE complex and Syt1, we used a single-vesicle content-mixing assay^{37,38} and tested the effect of both Syt1 and SNAP-25 mutants on association, spontaneous fusion and Ca^{2+} -triggered fusion of single vesicles with reconstituted full-length neuronal SNAREs, Syt1 and complexin-1 (Fig. 3d–g, Extended Data Fig. 8 and Extended Data Table 2). The Syt1 mutants disrupting the interface between the SNARE complex and Syt1 as seen by co-immunoprecipitation (Fig. 3b, c) also reduced vesicle association to similar degrees for the four different Syt1 mutants (left group of mutants in Fig. 3d). Our reconstituted assay also allowed testing of the interacting SNAP-25 residues (middle group of mutants in Fig. 3d), showing significant reduction of vesicle association as well. Thus, as expected, mutations on both the Syt1 and SNARE sides of the primary interface reduce the interaction between them. Likewise, both groups of mutants reduce the amplitude and decrease the synchronization of Ca^{2+} -triggered fusion, but they do not affect spontaneous fusion (Fig. 3e–g). In particular, the Syt1 quintuple mutant significantly reduced Ca^{2+} -triggered synchronization to the control level without Syt1.

To characterize the interaction between Syt1 and the SNARE complex further, we investigated the effect of ATP, which has an ionic shielding effect on certain cellular processes^{39,40}. We observed no effect on both spontaneous and Ca^{2+} -triggered fusion, and only a mild effect on vesicle association (Fig. 3d–g, right control group), suggesting that the functionally important interactions between Syt1 and the SNARE complex are not affected by ionic shielding.

Evoked release requires primary interface

We assayed release electrophysiologically in Syt1 conditional knockout neurons in which endogenous Syt1 was replaced with mutant Syt1 (Fig. 3a and Methods). As expected⁷, removal of Syt1 abolished synchronous release, as monitored by recording evoked inhibitory postsynaptic currents (eIPSCs), a phenotype that can be rescued by re-introduction of wild-type but not mutant Syt1 cDNA (Fig. 4a, b). Notably, the region I mutants retained some rescue ability, in contrast to region II mutants that were nearly non-functional (Fig. 4a, b), including the R398Q/R399Q mutant as previously reported³⁶. The triple mutant (R281A/R398A/R399A) and the Syt1 quintuple mutant that combines mutations in both regions showed even more severe phenotypes. In contrast, a mutant that affects the polybasic region of Syt1 (R322E/K325E) exhibited a milder phenotype, as reported

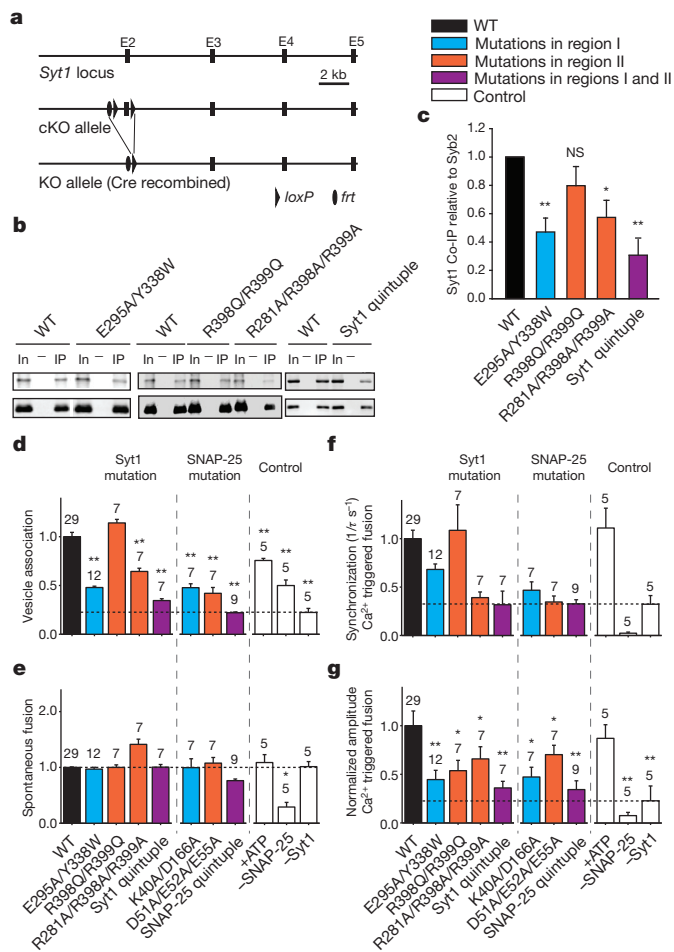


Figure 3 | Mutations of the primary interface affect binding and Ca^{2+} -triggered single vesicle-vesicle fusion. ‘Syt1 quintuple’ refers to the Syt1 mutant (R281A/E295A/Y338W/R398A/R399A). ‘SNAP-25 quintuple’ refers to the SNAP-25 mutant (K40A/D51A/E52A/E55A/D166A). The colour code is specified in the figure. **a**, Schematic diagram of Syt1 conditional knockout (cKO) mice. The *Syt1* exon 2 which contains the transmembrane domain is floxed. Cre recombinase removes exon 2, ablating all cytoplasmic *Syt1* sequences. **b**, Co-immunoprecipitation (Co-IP) of either Syt1 (top row) or synaptobrevin-2 (bottom row) with a syntaxin-1A antibody in Syt1 conditional knockout cultured neurons rescued with the indicated Syt1 mutant constructs. **c**, Quantification of co-immunoprecipitation of Syt1 normalized to synaptobrevin-2. Results are scaled to Syt1 wild-type levels. All data are means \pm s.e.m.; statistical significance was analysed by the Student’s *t*-test comparing the mutants with wild-type Syt1; $^{**}P < 0.01$, $n = 4$ for Syt1 E295A/Y338W and Syt1 quintuple; NS, no significant difference, $n = 3$ for Syt1 R398Q/R399Q; $^{*}P < 0.05$, $n = 4$ for Syt1 R281A/R398A/R399A. **d–g**, Bar graphs showing the effects of Syt1 and SNAP-25 mutants in fusion of single vesicles with reconstituted neuronal SNAREs, Syt1 and complexin-1 (see Methods and ref. 38). **d**, Number of associated SV vesicles (see Methods) after incubation of SV vesicles with surface-immobilized PM vesicles (see Methods) for a 1-min period. **e**, Number of spontaneous fusion events over the subsequent 1-min observation period normalized by the number of associated SV vesicles. **f**, Synchronization, that is, decay rates ($1/\tau$), of the histograms of fusion events upon $500 \mu\text{M}$ Ca^{2+} injection. Error bars are error estimates computed from the covariance matrix upon fitting the corresponding cumulative histograms with a single exponential decay function using a Levenberg–Marquardt technique. **g**, Amplitude of the first 1-s time bin upon Ca^{2+} injection. Each value in this panel was normalized by the respective number of fusion events after Ca^{2+} injection. **d, e, g**, All data are means \pm s.e.m.; the number of independent repeat experiments are depicted above the bars and in Extended Data Table 2; statistical significance was assessed by the Student’s *t*-test comparing all other conditions with wild type ($^{**}P < 0.005$; $^{*}P < 0.05$). The cumulative fusion histograms are shown in Extended Data Fig. 8. Controls in panels **d–g** are in the presence of 3 mM ATP and in the absence of SNAP-25 and Syt1. As expected, Ca^{2+} -triggered fusion required the presence of both SNAP-25 and Syt1; fusion is not affected by the presence of ATP.

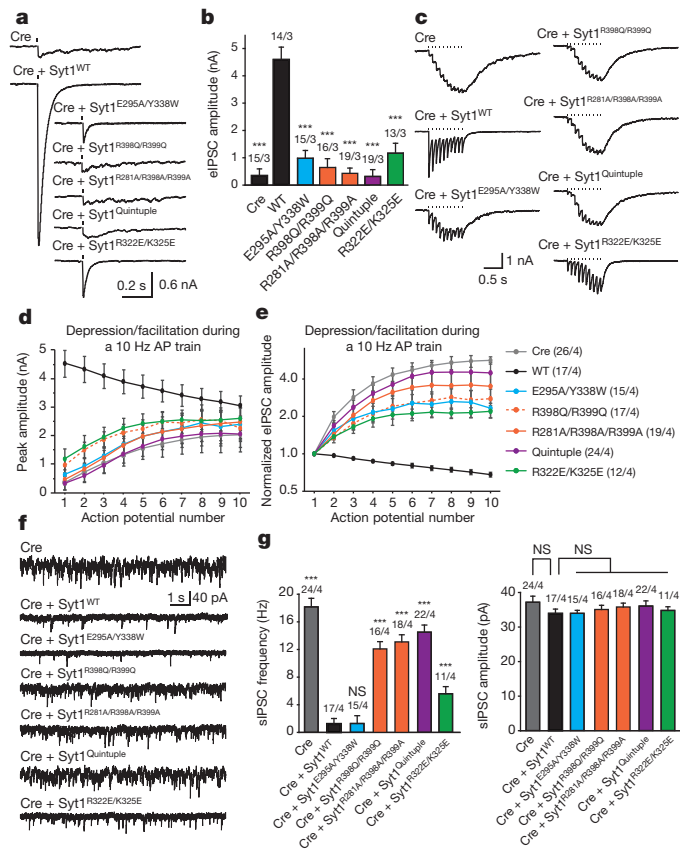


Figure 4 | Mutations of the primary interface impair Syt1 function in Ca²⁺-triggered release. Recording of inhibitory postsynaptic currents (IPSCs) from cultured Syt1 conditional knockout hippocampal neurons infected with lentiviruses expressing Cre recombinase and Syt1 mutants of the primary interface and a mutant of the polybasic region (R322E/K325E, ref. 30). All recordings were performed in the presence of 6-cyano-7-nitroquinoxaline-2,3-dione (CNQX, 20 μ M) and D-2-amino-5-phosphonovalerate (AP-5, 50 μ M) using a high Cl⁻ internal solution. **a, b**, Sample traces of evoked IPSCs from single action potentials (**a**) and quantification of peak amplitudes (**b**). Tick marks indicate stimulus delivery. All data are means \pm s.e.m.; number of cells/independent cultures analysed are depicted above the bars; statistical significance was assessed by one-way analysis of variance comparing all other conditions with wild-type rescue group ($***P < 0.001$). **c–e**, Syt1 mutants display facilitation, instead of depression, during high-frequency stimulation. **c**, Sample traces of 10 Hz trains; **d, e**, quantification of absolute (**d**) and normalized (**e**) IPSC amplitudes during the train; numbers of cells/independent cultures analysed are depicted in parentheses in the labels for each of the traces. AP, action potential. **f, g**, Syt1 mutants are unable to clamp the frequency of spontaneous IPSCs (sIPSCs). **f**, Sample spontaneous IPSC traces; **g**, quantification of event frequency (left) and amplitude (right). All data are means \pm s.e.m.; the number of cells/independent cultures analysed are depicted above the bars. Statistical significance was assessed by one-way analysis of variance comparing all other conditions with the wild-type rescue group ($***P < 0.001$; NS, no significant difference).

previously³⁰. During high-frequency stimulation, wild-type cultured hippocampal neurons displayed depression while Syt1 conditional knockout neurons showed asynchronous release with robust facilitation (Fig. 4c–e). All Syt1 mutants underwent facilitation, and the severity of the phenotype for each mutant correlated well with the results for single evoked release. Another known consequence of Syt1 removal is the unclamping of spontaneous release. Interestingly, the E295A/Y338W mutant could rescue this phenotype while the region II mutants (or combinations thereof) could not (Fig. 4f, g). There were no differences in spontaneous miniature IPSC amplitudes between wild type and mutant rescues (Fig. 4g).

Together, these observations support the notion that the primary Syt1-SNARE interface is critical for the role of Syt1 as a Ca^{2+} sensor

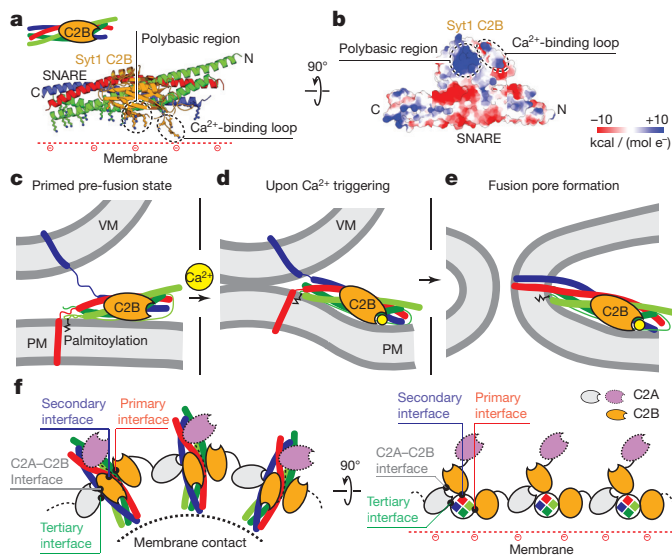


Figure 5 | Model of the role of the primary Syt1–SNARE interface. **a, b,** Unit that is formed by the primary interface between Syt1 C2B and the SNARE complex. **a,** Cartoon representation with positively charged side chains shown as sticks; **b,** electrostatic potential map looking towards the positively charged face of the Syt1 C2B–SNARE unit. **c–e,** Proposed function of the Syt1 C2B–SNARE unit. **c,** Initial state before Ca^{2+} triggering. The juxtamembrane linkers of synaptobrevin-2 and of syntaxin-1A were modelled as random coils³⁷. **d,** Intermediate state after Ca^{2+} triggering when the membranes are close enough to promote stalk formation⁴⁸. **e,** Fusion pore formation. Zigzag lines indicate palmitoylated cysteine residues of the SNAP-25 linker region. PM, plasma membrane; VM, vesicle membrane. **f,** Other interfaces found in the crystal structure, along with the primary interface, could form a connected network of SNARE complexes that surrounds the point of contact between membranes. The left panel is a top-down view onto the point of contact between membranes; the right panel is a rotated projection view.

of evoked release, but that this Syt1–SNARE interface is not required for spontaneous release in neurons, consistent with the differential effect of Syt1 mutations on spontaneous and evoked release^{11–13}. We note that spontaneous release observed in our reconstituted system³⁸ reflects the fusion probability at exactly zero Ca^{2+} concentration (that is, in the absence of a Ca^{2+} sensor for mini-release), whereas spontaneous release in neurons is driven by resting Ca^{2+} concentrations and is largely blocked by removing all calcium⁴¹. This circumstance may explain the different behaviour for the E295A/Y338W mutant in neuronal cultures and in the reconstituted system. Moreover, there is little effect on synchronization of the R398Q/R399Q mutant of Syt1 in the reconstituted system, which may indicate the existence of additional interactions involving these residues with other factors not present in our minimal reconstituted system.

Model of Syt1-mediated Ca^{2+} triggering

Several mechanisms and models have been proposed for how Syt1 and the neuronal SNARE complex cooperate^{29,42,43}. These models, however, were devised based on biochemical data without atomic-resolution structural information. Our crystal structures of the Syt1–SNARE complex now suggest a parsimonious mechanism that involves a concerted action of the unit consisting of the Syt1 C2B domain and the SNARE complex (Fig. 5a). We first focus on this unit alone and then later place it in the context of full-length Syt1 and other interactions. The combined surface of the Syt1 C2B–SNARE unit forms a flat face with an extensive pattern of positive charges (Fig. 5b) that includes the polybasic region of Syt1 as well as basic residues at the membrane-proximal side of the neuronal SNARE complex. The effect of the Syt1 R322E/K325E mutant (Fig. 4) as well as reported observations^{29,30,35,44,45} suggests a functional role for the polybasic region.

Before Ca^{2+} influx, we assume a membrane juxtaposed, hemifusion-free state (Fig. 5c), the likely starting state for fast Ca^{2+} -triggered fusion³⁷, and a partially folded *trans*-SNARE complex⁴⁶. We assume that the Syt1 C2B–SNARE unit can assemble in this state since the primary interface involves residues within the folded region of the *trans*-SNARE complex. The palmitoylated cysteine residues of SNAP-25 would be able to interact with the plasma membrane, and the membrane-proximal side of syntaxin is closely juxtaposed with the membrane, as inferred by the requirement of tight membrane coupling of syntaxin for evoked release⁴⁷. Upon Ca^{2+} binding to the Syt1 C2B domain, the Ca^{2+} binding loops partially insert into the membrane²³ with a preference for membrane curvature^{20,21}, possibly supported by ionic interactions via the polybasic region and with continuing maintenance of the Syt1–SNARE interface. We propose that the Syt1 C2B–SNARE unit moves en bloc as an entity upon Ca^{2+} triggering. The simultaneous membrane interactions of the Ca^{2+} binding loops, of the polybasic region, and of the membrane-proximal region of the SNARE complex would therefore require a deformation of the plasma membrane. This morphological change of the plasma membrane juxtaposes the membranes closer than the critical distance (0.9 nm) to promote stalk formation⁴⁸ (Fig. 5d), and subsequently leads to fusion pore opening (Fig. 5e).

It is likely that Syt1 engages in multiple functionally relevant interactions with neuronal SNARE complexes, including the other interfaces found in our crystal structures. All of these interactions may be anchored by the primary interface, which is the most stable and extensive contact site. We speculate that multiple SNARE complexes could interact via Syt1 interactions employing some or all of the interfaces (primary, secondary, tertiary, C2A–C2B) that are observed in our crystal structures (Fig. 5f). In addition to the primary and secondary interfaces formed by C2B domains (Fig. 5f, gold), the tertiary interface involves a C2A domain (Fig. 5f, grey). This interface could be involved in displacing complexin from its interaction with the core of the SNARE complex, since complexin binds in the groove between syntaxin-1A and synaptobrevin-2 (ref. 49); partial displacement of complexin may be important for neurotransmitter release⁵⁰. Interestingly, the gain-of-function mutation D232N in the Ca^{2+} -binding region of the C2A domain²⁶ is close to the tertiary interface and it is also part of the C2A–C2B interface (Fig. 1b). In the proposed network of interactions there are C2A domains (Fig. 5f, purple) that have no interactions with SNARE complexes; we envision that they could interact with a membrane. This entire Syt1–SNARE assembly would thus be poised to confer cooperativity upon Ca^{2+} triggering on a sub-millisecond timescale.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 23 March; accepted 27 July 2015.

Published online 17 August 2015.

- Wickner, W. & Schekman, R. Membrane fusion. *Nature Struct. Mol. Biol.* **15**, 658–664 (2008).
- Rothman, J. E. The Principle of Membrane Fusion in the Cell (Nobel Lecture). *Angew. Chemie Int. Ed.* **53**, 12676–12694 (2014).
- Sutton, R. B., Fasshauer, D., Jahn, R. & Brunger, A. T. Crystal structure of a SNARE complex involved in synaptic exocytosis at 2.4 Å resolution. *Nature* **395**, 347–353 (1998).
- Weber, T. et al. SNAREpins: minimal machinery for membrane fusion. *Cell* **92**, 759–772 (1998).
- Südhof, T. C. Neurotransmitter release: the last millisecond in the life of a synaptic vesicle. *Neuron* **80**, 675–690 (2013).
- Pang, Z. P. & Südhof, T. C. Cell biology of Ca^{2+} -triggered exocytosis. *Curr. Opin. Cell Biol.* **22**, 496–505 (2010).
- Geppert, M. et al. Synaptotagmin I: a major Ca^{2+} sensor for transmitter release at a central synapse. *Cell* **79**, 717–727 (1994).
- Xu, J., Mashimo, T. & Südhof, T. C. Synaptotagmin-1, -2, and -9: Ca^{2+} sensors for fast release that specify distinct presynaptic properties in subsets of neurons. *Neuron* **54**, 567–581 (2007).
- Wen, H. et al. Distinct roles for two synaptotagmin isoforms in synchronous and asynchronous transmitter release at zebrafish neuromuscular junction. *Proc. Natl Acad. Sci. USA* **107**, 13906–13911 (2010).

10. Bacaj, T. *et al.* Synaptotagmin-1 and synaptotagmin-7 trigger synchronous and asynchronous phases of neurotransmitter release. *Neuron* **80**, 947–959 (2013).
11. Yoshihara, M. & Littleton, J. T. Synaptotagmin I functions as a calcium sensor to synchronize neurotransmitter release. *Neuron* **36**, 897–908 (2002).
12. Maximov, A. & Südhof, T. C. Autonomous function of synaptotagmin 1 in triggering synchronous release independent of asynchronous release. *Neuron* **48**, 547–554 (2005).
13. Kochubei, O. & Schneggenburger, R. Synaptotagmin increases the dynamic range of synapses by driving Ca^{2+} -evoked release and by clamping a near-linear remaining Ca^{2+} sensor. *Neuron* **69**, 736–748 (2011).
14. Davletov, B. A. & Südhof, T. C. A single C2 domain from synaptotagmin I is sufficient for high affinity Ca^{2+} /phospholipid binding. *J. Biol. Chem.* **268**, 26386–26390 (1993).
15. Brose, N., Petrenko, A. G., Südhof, T. C. & Jahn, R. Synaptotagmin: a calcium sensor on the synaptic vesicle surface. *Science* **256**, 1021–1025 (1992).
16. Fernández-Chacón, R. *et al.* Synaptotagmin I functions as a calcium regulator of release probability. *Nature* **410**, 41–49 (2001).
17. Vrljic, M. *et al.* Post-translational modifications and lipid binding profile of insect cell-expressed full-length mammalian synaptotagmin 1. *Biochemistry* **50**, 9998–10012 (2011).
18. Rhee, J.-S. *et al.* Augmenting neurotransmitter release by enhancing the apparent Ca^{2+} affinity of synaptotagmin 1. *Proc. Natl Acad. Sci. USA* **102**, 18664–18669 (2005).
19. Araç, D. *et al.* Close membrane-membrane proximity induced by Ca^{2+} -dependent multivalent binding of synaptotagmin-1 to phospholipids. *Nature Struct. Mol. Biol.* **13**, 209–217 (2006).
20. Hui, E., Johnson, C. P., Yao, J., Dunning, F. M. & Chapman, E. R. Synaptotagmin-mediated bending of the target membrane is a critical step in Ca^{2+} -regulated fusion. *Cell* **138**, 709–721 (2009).
21. Martens, S., Kozlov, M. M. & McMahon, H. T. How synaptotagmin promotes membrane fusion. *Science* **316**, 1205–1208 (2007).
22. Zhang, X., Rizo, J. & Südhof, T. C. Mechanism of phospholipid binding by the C2A domain of synaptotagmin I. *Biochemistry* **37**, 12395–12403 (1998).
23. Rufener, E., Frazier, A. A., Wieser, C. M., Hinderliter, A. & Cafiso, D. S. Membrane-bound orientation and position of the synaptotagmin C2B domain determined by site-directed spin labeling. *Biochemistry* **44**, 18–28 (2005).
24. Bennett, M. K., Calakos, N. & Scheller, R. H. Syntaxin: a synaptic protein implicated in docking of synaptic vesicles at presynaptic active zones. *Science* **257**, 255–259 (1992).
25. Li, C. *et al.* Ca^{2+} -dependent and -independent activities of neural and non-neural synaptotagmins. *Nature* **375**, 594–599 (1995).
26. Pang, Z. P., Shin, O.-H., Meyer, A. C., Rosenmund, C. & Südhof, T. C. A gain-of-function mutation in synaptotagmin-1 reveals a critical role of Ca^{2+} -dependent soluble N-ethylmaleimide-sensitive factor attachment protein receptor complex binding in synaptic exocytosis. *J. Neurosci.* **26**, 12556–12565 (2006).
27. Rickman, C. *et al.* Conserved prefusion protein assembly in regulated exocytosis. *Mol. Biol. Cell* **17**, 283–294 (2006).
28. Choi, U. B. *et al.* Single-molecule FRET-derived model of the synaptotagmin 1-SNARE fusion complex. *Nature Struct. Mol. Biol.* **17**, 318–324 (2010).
29. Dai, H., Shen, N., Araç, D. & Rizo, J. A quaternary SNARE-synaptotagmin- Ca^{2+} -phospholipid complex in neurotransmitter release. *J. Mol. Biol.* **367**, 848–863 (2007).
30. Brewer, K. D. *et al.* Dynamic binding mode of a Synaptotagmin-1–SNARE complex in solution. *Nature Struct. Mol. Biol.* **22**, 555–564 (2015).
31. Sutton, R. B., Davletov, B. A., Berghuis, A. M., Südhof, T. C. & Sprang, S. R. Structure of the first C2 domain of synaptotagmin I: a novel Ca^{2+} /phospholipid-binding fold. *Cell* **80**, 929–938 (1995).
32. Fernandez, I. *et al.* Three-dimensional structure of the synaptotagmin 1 C2B domain: synaptotagmin 1 as a phospholipid binding machine. *Neuron* **32**, 1057–1069 (2001).
33. Fuson, K. L., Montes, M., Robert, J. J. & Sutton, R. B. Structure of human synaptotagmin 1 C2AB in the absence of Ca^{2+} reveals a novel domain association. *Biochemistry* **46**, 13041–13048 (2007).
34. Redecke, L. *et al.* Natively inhibited *Trypanosoma brucei* cathepsin B structure determined by using an X-ray laser. *Science* **339**, 227–230 (2013).
35. Gaffaney, J. D., Dunning, F. M., Wang, Z., Hui, E. & Chapman, E. R. Synaptotagmin C2B domain regulates Ca^{2+} -triggered fusion *in vitro*: critical residues revealed by scanning alanine mutagenesis. *J. Biol. Chem.* **283**, 31763–31775 (2008).
36. Xue, M., Ma, C., Craig, T. K., Rosenmund, C. & Rizo, J. The Janus-faced nature of the C2B domain is fundamental for synaptotagmin-1 function. *Nature Struct. Mol. Biol.* **15**, 1160–1168 (2008).
37. Diao, J. *et al.* Synaptic proteins promote calcium-triggered fast transition from point contact to full fusion. *Elife* **1**, e00109 (2012).
38. Lai, Y. *et al.* Complexin inhibits spontaneous release and synchronizes Ca^{2+} -triggered synaptic vesicle fusion by distinct mechanisms. *Elife* **3**, 1–14 (2014).
39. Suh, B.-C. & Hille, B. Electrostatic interaction of internal Mg^{2+} with membrane PIP₂ Seen with KCNQ K⁺ channels. *J. Gen. Physiol.* **130**, 241–256 (2007).
40. Park, Y. *et al.* Controlling synaptotagmin activity by electrostatic screening. *Nature Struct. Mol. Biol.* **19**, 991–997 (2012).
41. Xu, J., Pang, Z. P., Shin, O.-H. & Südhof, T. C. Synaptotagmin-1 functions as a Ca^{2+} sensor for spontaneous release. *Nature Neurosci.* **12**, 759–766 (2009).
42. Krishnakumar, S. S. *et al.* Conformational dynamics of calcium-triggered activation of fusion by synaptotagmin. *Biophys. J.* **105**, 2507–2516 (2013).
43. Wang, J. *et al.* Calcium sensitive ring-like oligomers formed by synaptotagmin. *Proc. Natl Acad. Sci. USA* **111**, 13966–13971 (2014).
44. Masumoto, T. *et al.* Ca^{2+} -independent syntaxin binding to the C2B effector region of synaptotagmin. *Mol. Cell. Neurosci.* **49**, 1–8 (2012).
45. Mackler, J. & Reist, N. Mutations in the second C2 domain of synaptotagmin disrupt synaptic transmission at *Drosophila* neuromuscular junctions. *J. Comp. Neurol.* **436**, 4–16 (2001).
46. Gao, Y. *et al.* Single reconstituted neuronal SNARE complexes zipper in three distinct stages. *Science* **337**, 1340–1343 (2012).
47. Zhou, P., Bacaj, T., Yang, X., Pang, Z. P. & Südhof, T. C. Lipid-anchored SNAREs lacking transmembrane regions fully support membrane fusion during neurotransmitter release. *Neuron* **80**, 470–483 (2013).
48. Aeffer, S., Reusch, T., Weinhausen, B. & Salditt, T. Energetics of stalk intermediates in membrane fusion are controlled by lipid composition. *Proc. Natl Acad. Sci. USA* **109**, E1609–E1618 (2012).
49. Chen, X. *et al.* Three-dimensional structure of the complexin/SNARE complex. *Neuron* **33**, 397–409 (2002).
50. Tang, J. *et al.* A complexin/synaptotagmin 1 switch controls fast synaptic vesicle exocytosis. *Cell* **126**, 1175–1187 (2006).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank M. S. Padolina for help with protein purification, the Northeastern Collaborative Access Team (supported by NIH P41 GM103403) at Advanced Photon Source for X-ray data collection, and the SSRL/LCLS scientists E. L. Baxter, P. Ehrensberger, T. I. Eriksson, Y. Feng, M. Hollenbeck, E. G. Kovaleva, S. E. McPhillips, S. Nelson, J. Song, Y. Tsai, V. Vinetsky and D. Zhu for their invaluable assistance with data collection at the LCLS XPP facility. Use of the Stanford Synchrotron Radiation Lightsource (SSRL) and Linac Coherent Light Source (LCLS), SLAC National Accelerator Laboratory, is supported by the US Department of Energy, Office of Science, Office of Basic Energy Sciences under contract no. DE-AC02-76SF00515. The SSRL Structural Molecular Biology Program is supported by the DOE Office of Biological and Environmental Research, and by the National Institutes of Health, National Institute of General Medical Sciences (including P41GM103393). Portions of this research were carried out at the Linac Coherent Light Source (LCLS) at the SLAC National Accelerator Laboratory. LCLS is an Office of Science User Facility operated for the US Department of Energy Office of Science by Stanford University. This research was supported in part by the National Institutes of Health (R37MH63105 to A.T.B.; MH086403 to T.C.S.; GM095887 and GM102520 to N.K.S. and A.S.B.); and by a HHMI Collaborative Innovation Award (HCIA) to A.T.B. and W.I.W.

Author Contributions Q.Z. designed, expressed, purified and crystallized the Syt1–SNARE complexes and the Syt1 C2B quintuple mutant, performed CD experiments, and designed the mutants to disrupt the primary interface. Q.Z. and M.Z. collected all diffraction data, determined and refined the crystal structures. Y.L. and J.D. performed the reconstituted single vesicle-vesicle experiments. T.B. performed the co-immunoprecipitation and electrophysiological experiments of neuronal cultures. A.Y.L., M.U., O.B.Z., N.K.S., A.S.B., W.I.W. and A.T.B. analysed and processed the LCLS diffraction data. A.E.C. and S.M.S. designed the goniometer-based setup at LCLS-XPP and helped with data collection. R.A.-M., M.C. and H.T.L. helped with data collection at LCLS-XPP. R.A.P. and Y.L. expressed and purified proteins for the single vesicle-vesicle experiments. U.B.C. helped with the comparison between the crystal structure and the smFRET data. Q.Z., M.Z., Y.L., T.B., T.C.S. and A.T.B. wrote the manuscript.

Author Information The coordinates of the atomic models and corresponding structure factors of the Syt1–SNARE complexes have been deposited in the Protein Data Bank under the accession codes 5CCG, 5CCH, 5CCI and 5CCJ. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.T.B. (brunger@stanford.edu) or T.C.S. (tcs1@stanford.edu).

METHODS

No statistical methods were used to predetermine sample size.

Strategy for crystallization of the Syt1–SNARE complex. Successful crystallization of the Syt1–SNARE complex involved extensive testing of multiple designs of covalently linked chimaeras between a Syt1 C2AB fragment and different SNARE domain fragments of the neuronal SNARE complex (both SNARE domains of SNAP-25, denoted SNAP-25_N and SNAP-25_C, as well as the SNARE domain of syntaxin-1A). Three different linker lengths for the chimaeras were tested (16, 23 and 37 amino acids) derived from the linker sequence of the human Oct-1 transcription factor⁵¹. This particular linker sequence had been used previously to crystallize the Arf–ArfGAP complex⁵². We also tested different truncations of the neuronal SNARE complex^{3,53}. The resulting constructs were screened for protein expression and homogeneity by ion exchange and size-exclusion chromatography. The best constructs resulted in a Syt1–SNARE complex that eluted in a mono-disperse peak in the final size-exclusion chromatography step, and that remained stable in SDS–PAGE without boiling, a hallmark for neuronal SNARE complex formation (Extended Data Fig. 1b). Best results were obtained when the SNARE fragments were truncated as previously described⁵³. The best candidates were used for crystallization trials (see below).

Cloning, expression and purification of the Syt1–SNARE complex. The SNAP-25 isoform used throughout this study is commonly referred to as isoform 2 or SNAP-25A. The C2AB fragment of rat synaptotagmin-1 (amino acid range 141–421) was fused to the amino terminus of the C-terminal SNARE domain of rat SNAP-25 (SNAP-25_C, amino acid range 141–204) via a 37-amino-acid linker (sequence NLSSDSSLSPSALNSLSSPSALNSTASNSPGIEGLS) derived from the human Oct-1 transcription factor⁵¹ (Extended Data Fig. 1a) (referred to as C2AB-linker-SNAP-25_C).

The C2AB-linker-SNAP-25_C chimaera, the rat SNAP-25_N fragment (amino acid range 7–83), the rat syntaxin-1A fragment (amino acid range 191–256), and the His-tagged rat synaptobrevin-2 fragment (amino acid range 28–89) were cloned into the Duet expression system (Novagen) following previous work with the neuronal SNARE complex⁵⁴ (Extended Data Fig. 1a). These four protein constructs were co-expressed in *Escherichia coli*, leading to complex formation in the host (referred to as Syt1–SNARE^{37aa-linker} complex). Specifically, *E. coli* BL21(DE3) cells were grown overnight at 30 °C using auto-inducing LB medium⁵⁵. After harvesting the cells by centrifugation, the pellet was re-suspended in lysis buffer (50 mM Tris-HCl, pH 8.0, 300 mM NaCl, 20 mM imidazole, 1 mM CaCl₂, 0.5 mM TCEP and EDTA-free protease inhibitor cocktail (Roche)) and lysed by three passes through the Emulsiflex C5 homogenizer (Avestin) at 15,000 p.s.i. After centrifugation, the cleared lysate was bound to a 4-ml bed volume of Ni-NTA beads (Qiagen) equilibrated in the lysis buffer. Beads were harvested by centrifugation and poured into a column, washed with the lysis buffer, and subsequently washed with the lysis buffer supplemented with additional 40 mM imidazole. The Syt1–SNARE^{37aa-linker} complex was eluted with the lysis buffer supplemented with additional 350 mM imidazole.

Depending on purification of Ca²⁺-free or Ca²⁺-bound complex, EDTA or CaCl₂ was included at specified steps. The fresh eluent of the Ni-NTA-affinity purified Syt1–SNARE^{37aa-linker} complex was pooled and dialysed against dialysis buffer I (50 mM Tris-HCl, pH 8.0, 150 mM NaCl, 0.5 mM TCEP; with 1 mM CaCl₂ or 10 mM EDTA) for 3 to 4 h at 4 °C. The dialysate was supplemented with TEV protease, and further dialysed in dialysis buffer II (50 mM Tris-HCl, pH 8.0, 50 mM NaCl, 0.5 mM TCEP; with 1 mM CaCl₂ or none) overnight at 4 °C. After removal of uncleaved sample, the His-tag-cleaved complex was subjected to anion exchange chromatography (buffer A: 50 mM Tris-HCl, pH 8.0, 50 mM NaCl, 0.5 mM TCEP, buffer B: 50 mM Tris-HCl, pH 8.0, 500 mM NaCl, 0.5 mM TCEP; both buffer A and B were supplemented with 1 mM CaCl₂ or 2 mM EDTA) using a linear gradient of NaCl starting at 50 mM and ending at 250 mM. The peak fractions were pooled, concentrated and loaded onto a Superdex 200 10/300 GL column (GE Healthcare) that was pre-equilibrated with SEC buffer (20 mM Tris-HCl, pH 8.0, 300 mM NaCl, 0.5 mM TCEP) for Ca²⁺-free complex, and SEC buffer supplemented with 1 mM CaCl₂ for Ca²⁺-bound complex (Extended Data Fig. 1b). The peak fractions containing Syt1–SNARE^{37aa-linker} complex were pooled and concentrated to a final concentration of ~20 mg ml⁻¹ for crystallization.

Cloning, expression and purification of wild-type and mutant Syt1 C2B domains. All wild-type and mutant synaptotagmin-1 C2B domains were prepared by using a standard PCR-based protocol. All PCR products were subcloned into the pGEX-6P-1 (GE Healthcare) and expressed as GST-tagged fusion proteins in *E. coli* BL21(DE3) cells at 30 °C overnight using auto-inducing LB medium⁵⁵. After harvesting the cells by centrifugation, the sample was resuspended in lysis buffer containing 50 mM HEPES-Na, pH 7.5, 300 mM NaCl, 2 mM DTT and EDTA-free protease inhibitor cocktail (Roche), and then subjected to sonication and centrifugation. The supernatant was incubated with glutathione-sepharose

beads (GE Healthcare). The resin was extensively washed with 50 ml of wash buffer I containing 50 mM HEPES-Na, pH 7.5, 300 mM NaCl, and 1 mM DDT, followed by 50 ml of wash buffer II containing 50 mM HEPES-Na, pH 7.5, 300 mM NaCl, 1 mM DTT, and 50 mM CaCl₂. The GST tag was cleaved overnight at 4 °C with PreScission protease (GE Healthcare) in cleavage buffer containing 50 mM HEPES-Na, pH 7.5, 300 mM NaCl, 1 mM DTT, 2 mM EDTA. The cleaved proteins were purified by gel filtration on Superdex 75 (GE Healthcare) that was pre-equilibrated with SEC buffer containing 20 mM Tris-HCl, pH 8.0, 150 mM NaCl, 0.5 mM TCEP. The peak fractions were pooled and concentrated to a final concentration of ~30 mg ml⁻¹ for circular dichroism analysis and crystallization.

Crystallization of the Ca²⁺-bound Syt1–SNARE complex. Crystallization trials performed by mixing the neuronal SNARE complex and Syt1 C2AB fragments were unsuccessful, probably due to the weak affinity between the two components and the aggregation propensity at higher concentrations, especially in the presence of Ca²⁺. To overcome these problems, crystallization trials were performed with chimaeras between Syt1 C2AB and SNAP-25_C, as described above. The best three candidates consisted of three different linker lengths. Initially, thin crystalline plates that diffracted to about 30 Å were obtained from a construct connected by a 23-amino-acid linker. Screening of volatile crystallization additives (such as, methanol, ethanol, 1,2-butanediol) led to improved diffraction to about 10 Å resolution, with sharp Bragg peaks. The complex with a longer 37-amino-acid linker (Syt1–SNARE^{37aa-linker}) was more soluble in SEC buffer compared to the one with the shorter 23-amino-acid linker. Clusters of needle-shaped crystals were obtained for this construct at first. Using a reverse vapour-diffusion method led to thicker crystal plates (Extended Data Fig. 1e), and eventually produced Bragg reflections past 4 Å resolution. There were consistently two crystal forms in the same drop with identical morphology, referred to as ‘short unit cell’ crystal form and ‘long unit cell’ crystal form, respectively. The difference between the short and long unit cell crystal forms of the Ca²⁺-bound Syt1–SNARE complex can be approximately described by a doubling of the number of complexes, except that one of the interacting Syt1 C2AB fragments is absent (Fig. 1c, Extended Data Fig. 2g and Extended Data Fig. 3d).

Before setting up crystal trays, purified Syt1–SNARE^{37aa-linker} complex (at a concentration of ~20 mg ml⁻¹) was diluted to a final concentration of ~8 mg ml⁻¹ supplemented with CaCl₂ and MgCl₂. The final buffer contained 20 mM Tris-HCl (pH 8.0), 300 mM NaCl, 100 mM MgCl₂, 1 mM CaCl₂ and 0.5 mM TCEP. The reservoir contained 100 mM HEPES-Na (pH 7.5) and 1% PEG 8000. Equal amounts of protein and reservoir (2 µl) were mixed and incubated at 20 °C. The crystals appeared after 1 to 4 months, and were flash-frozen in a cryo-protecting solution containing the same constituents as the crystallization condition supplemented with 35% (v/v) sucrose.

Although Syt1 was initially covalently linked to the C-terminal half of SNAP-25 (SNAP-25_C), the linker was slowly cleaved at ambient temperature (Extended Data Fig. 1c). Moreover, crystal growth required 1 to 4 months and contained entirely cleaved complex (Extended Data Fig. 1b), enabling formation of the 2:1 and 3:2 stoichiometries in the asymmetric units of the short and long unit cell crystal forms, respectively. Thus, the crystal structures are probably not affected by the initial presence of the linker. We simply refer to the resulting crystal structure as that of the Syt1–SNARE complex.

Crystallization of the Mg²⁺-bound Syt1–SNARE complex. Crystals of Mg²⁺-bound Syt1–SNARE complex were grown using the reverse hanging-drop vapour-diffusion method. Purified Syt1–SNARE^{37aa-linker} (at a concentration of ~20 mg ml⁻¹) that was prepared without Ca²⁺ and in the presence of EDTA (as described above) was diluted to a final concentration of ~4 mg ml⁻¹ and supplemented with 100 mM MgCl₂. The final buffer contained 20 mM Tris-HCl (pH 8.0), 300 mM NaCl, 100 mM MgCl₂, and 0.5 mM TCEP. The reservoir contained 100 mM HEPES-Na (pH 7.5) and 2.5% PEG3350. Similar to Ca²⁺-bound complex, the linker was also cleaved during crystallization. Mg²⁺-bound Syt1–SNARE complex crystals appeared after 2 months and were flash-frozen in a cryo-protecting solution containing the same constituents as the crystallization condition supplemented with 20% (v/v) glycerol.

Crystallization of the quintuple mutant of the Syt1 C2B domain. Crystals of the quintuple mutant (R281A/E295A/Y338W/R398A/R399A) of the Syt1 C2B domain were grown by the hanging-drop vapour diffusion method at 20 °C by mixing 2 µl protein solution (at a concentration of ~10 mg ml⁻¹) in a buffer containing 20 mM Tris-HCl, pH 8.0, 150 mM NaCl, and 0.5 mM TCEP) with equal volume of reservoir solution containing 100 mM Tris-HCl (pH 8.5), 1.5 M ammonium sulfate.

XFEL data collection and processing. The diffraction data of crystals of the long unit cell crystal form of the Ca²⁺-bound Syt1–SNARE complex (Extended Data Table 1) were collected at the X-ray Pump Probe (XPP) endstation using the LCLS

XFEL at the SLAC National Accelerator Laboratory at Stanford University equipped with a Rayonix MX325 X-ray imaging detector. The XFEL beam was focused to 30 μm at a nominal energy of 9 keV and a pulse duration of 40 fs in self-amplified stimulated emission (SASE) mode^{56,57}. Each 40-fs XFEL pulse at the XPP endstation at LCLS delivers 10^{12} photons, exceeding a dose of 30 MGy. The dose is so large that the diffraction volume is vaporized after exposure to a single XFEL pulse (Extended Data Fig. 2b). Nonetheless, the 40-fs laser pulse enables the so-called 'diffraction before destruction' data collection^{58,59}, whereby the diffraction pattern is recorded before the sample is destroyed.

The diffraction images were obtained from 148 frozen and cryo-protected crystals mounted in conventional cryo-loops at 100 K using a goniometer-based fixed-target sample delivery station⁶⁰. XFEL diffraction tests of hydrated crystals at room temperature did not show any marked improvement in comparison to cryo-preserved crystals, at least in terms of visible limiting resolution of the observed diffraction pattern. Furthermore, the entire crystal was damaged when exposed at room temperature, requiring many more crystals and much more time to collect a complete diffraction data set. As beam time and sample volume were limited, we decided to collect diffraction data at cryogenic temperature, since it was possible to obtain multiple diffraction images from the same frozen crystal.

In most cases, 2–20 diffraction images (depending on crystal size and quality) were obtained from a single crystal. Since the diffracting volume was destroyed by the XFEL beam, a different volume had to be exposed for every shot (Extended Data Fig. 2b). The crystal was translated at least 100 μm between exposures. During data collection we observed that improved diffraction quality could be achieved by placing consecutive shots far apart (that is, across the length of the crystal from one another), then later 'filling' between the first and second shots in the same 'shuttling' manner. Finally, care was taken to collect exposures at a variety of spindle (ϕ) angles to maximize the completeness of the final data set.

Of the 148 crystals used in the experiment, 113 crystals produced 578 images that could be processed; the other crystals did not diffract or showed multiple lattices. The long unit cell form occurred much more frequently (about 80%). Images indexed in the short unit cell crystal form were identified using a hierarchical clustering method⁶¹ and were omitted from further processing since it was not possible to obtain a complete data set in this crystal form. Other images were rejected during data processing as described below. The 309 diffraction images (from 72 crystals) in the final selection for the long unit cell crystal form were indexed and integrated with the cctbx.xfel suite of data-processing software^{62–64}, with the diffraction data processing parameters optimized by a grid search procedure (Lyubimov *et al.*, manuscript in preparation). The integrated diffraction data were subsequently scaled, merged and post-refined with the PRIME software⁶⁵.

We optimized the data integration process using a combination of mosaic quality analysis⁶⁴, highest number of bright reflections yielded by integration, and overlay of observed reflections on the actual integrated areas to determine whether nearly all visible reflections were integrated, and no unobserved reflections were predicted and integrated. We inspected overlays of diffraction images and predicted reflection positions to fine-tune the computational approaches and optimize the data processing of the XFEL data. The iterative scaling and post-refinement approach used by PRIME allowed the construction of a complete diffraction data set from the relatively small number of diffraction images.

Synchrotron beamline data collection. All other diffraction data sets (Extended Data Table 1) were collected using beamline 24ID-C of the Advanced Photon Source (APS) at Argonne National Laboratory (Argonne, IL). Diffraction data of the best crystals of both the Ca^{2+} -bound short unit cell crystal form and the Mg^{2+} -bound short unit cell crystal form of the Syt1-SNARE complex were indexed and integrated using the XDS software⁶⁶, and scaled and merged using the SCALA program in CCP4 package⁶⁷. Diffraction data of the quintuple mutant of the Syt1 C2B domain were indexed, integrated, scaled and merged using the XDS software⁶⁶.

Structure determination. The phases for all crystal structures of Syt1-SNARE complex were determined by molecular replacement with Phaser⁶⁸ using the rat SNARE complex (Protein Data Bank (PDB) code 1N7S), the rat Syt1 C2A domain (PDB code 3F04), and the rat Syt1 C2B domain (PDB code 1UOW) as search models. The structures were iteratively rebuilt and refined using the programs Coot⁶⁹, CNS1.3^{70,71}, and phenix.refine⁷² (Extended Data Table 1). Both the Ca^{2+} - and Mg^{2+} -bound Syt1-SNARE complex in the short unit cell crystal form were refined with phenix.refine⁷² using non-crystallographic symmetry (NCS) restraints, secondary structure restraints, and grouped *B*-factor refinement. The long unit cell crystal form of the Ca^{2+} -bound Syt1-SNARE complex was initially refined using CNS 1.3^{70,71}, with DEN restraints⁷³, restrained grouped *B*-factors and NCS restraints, then further refined with phenix.refine⁷² using non-crystallographic symmetry (NCS) restraints, secondary structure restraints,

and individual *B*-factor refinement. $mF_o - DF_c$ annealed omit maps (Extended Data Fig. 2c and Extended Data Fig. 3b) were calculated using phenix.refine⁷² using omit refinement consisting of three cycles of simulated annealing and grouped/individual *B*-factor refinement (grouped *B*-factor refinement for the Ca^{2+} -bound Syt1-SNARE complex in the short unit cell crystal form, individual *B*-factor refinement for the Ca^{2+} -bound Syt1-SNARE complex in the long unit cell form) with non-crystallographic symmetry (NCS) restraints and secondary structure restraints; residues 335–340 in Syt1 and 159–166 in SNAP-25 were omitted (that is, residues in region I of the primary interface between Syt1 and the SNARE complex), atoms within a 4 Å cushion around these omitted residues were kept fixed. The *R* values of the refined structures (Extended Data Table 1) are well within the range that is typical at the corresponding resolutions⁷⁴. Ramachandran analysis with MolProbity⁷⁵ indicated that 98% of the residues are in the favoured regions and none is in disallowed regions for both the Ca^{2+} - and Mg^{2+} -bound Syt1-SNARE complexes in the short unit cell crystal form, and that 97% of the residues are in the favoured regions and none is in disallowed regions for the Ca^{2+} -bound Syt1-SNARE complex in the long unit cell crystal form. The quality indicators for the crystal structures of the Syt1-SNARE complex are well within acceptable ranges indicated by the 'polygon' plot⁷⁵ produced by phenix.refine⁷² and by the validation report of the deposited structures and diffraction data.

The phases for crystal structure of the quintuple mutant of the Syt1 C2B domain were determined by molecular replacement with Phaser⁶⁸ using the rat Syt1 C2B domain (PDB code 1UOW) as the search model. The structure was iteratively built and refined using the program Coot⁶⁹, and phenix.refine⁷² (Extended Data Table 1). The final model consists of four Syt1 C2B molecules in the asymmetric unit. Ramachandran analysis with MolProbity⁷⁵ indicated that 98% of the residues are in the favoured regions and none are in disallowed regions.

The LCLS XFEL produced superior electron density maps. The diffraction data obtained at the LCLS XFEL extended to substantially higher resolution than data collected at the APS NE-CAT microfocus synchrotron beamline from similar crystals of the long cell form (Extended Data Fig. 2a, b). It is notable that only 1 out of 85 screened crystals in the long unit cell form diffracted to 4.1 Å at APS, while 61 out of ~72 long unit cell crystals diffracted to at least 3.5 Å at LCLS. In fact, only a lack of available XFEL beam time prevented us from collecting a complete diffraction data set beyond 3.5 Å. We note that similar improvements in limiting resolution of XFEL versus synchrotron diffraction images have been observed for GPCRs⁶⁰. Interestingly, while the short unit cell crystal form produced a limiting resolution of 3.6 Å at APS NE-CAT, the density maps are substantially less well defined for side chains compared to the LCLS data set (compare Extended Data Fig. 2c–f and Fig. 3b, c).

Taken together, the LCLS diffraction data set proved notably superior to the particular synchrotron-derived data sets that we collected in terms of limiting resolution and quality of the electron density maps, and was thus essential for more accurately determining side-chain positions. Moreover, in the LCLS XFEL crystal structure there is clear electron density for 19 Ca^{2+} bound to the Ca^{2+} binding sites of the Syt1 C2 domains and to a few additional sites on the surfaces of Syt1 and SNARE molecules (Extended Data Table 1). In contrast, electron density could be identified for only 7 Ca^{2+} in the short unit cell crystal form collected at the APS synchrotron. Taking into account the smaller number of molecules in the asymmetric unit of the short unit cell crystal form (two Syt1 and one SNARE complex compared to three Syt1 and two SNARE complexes), there are still fewer Ca^{2+} sites that were observed in electron density maps derived from the data set collected at APS, suggesting that Ca^{2+} -binding sites have been affected by radiation damage.

At present, it is difficult to assess the relative quality of XFEL diffraction data studied here with conventional rotation diffraction data measured at a synchrotron. We suspect that the standard diffraction data statistics (such as the merging *R* values) of rotation data are better due to the ability to directly measure full reflections (at least by summation of partials) without modelling partiality, which is still a relatively crude process even with the latest post-refinement approaches. In our opinion, these apparently poorer statistics obtained from the current state-of-the-art XFEL diffraction data processing methods are more than offset by the improved quality of the resulting electron density maps for diffraction data of crystals collected at the XFEL (Fig. 2c, d and Extended Data Fig. 2c–f). These maps, superior to those of the short unit cell data crystal structure collected at the APS NE-CAT microfocus synchrotron beamline at comparable limiting resolution (Extended Data Fig. 3b, c), enabled us to better investigate the binding interfaces between Syt1 and the SNARE complex, and was thus essential for the structural portion of our study.

Validation and structure analysis. MolProbity⁷⁵ was used for evaluating the geometry and quality of the models (Extended Data Table 1). The electrostatic

potential maps were calculated and displayed using the UCSF Chimera package⁷⁶ (Chimera is developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco (supported by NIGMS P41-GM103311)). All other structure figures were prepared with PyMol (DeLano, 2002, The PyMOL Molecular Graphics System, <http://www.pymol.org>, Schrödinger, LLC). Interface areas were calculated by PISA⁷⁷; note that the commonly used 'buried surface area' is twice the 'interface area'.

Design of mutations to disrupt the primary interface. We mutated combinations of region I interacting residues (Syt1 E295A/Y338W and SNAP-25 K40A/D166A), and region II interacting residues (Syt1 R281A/R398A/R399A and SNAP-25 D51A/E52A/E55A). We verified that the Syt1 mutants are properly folded by measuring circular dichroism (CD) spectra and thermal denaturation curves (Extended Data Fig. 7a), crystal structure determination of the Syt1 C2B quintuple mutant (R281A/E295A/Y338W/R398A/R399A) (Extended Data Fig. 7b–d and Extended Data Table 1), and size-exclusion chromatography of Syt1 and its mutants (Extended Data Fig. 7e). The crystal structure of the quintuple mutant is very similar to that of wild-type C2B (PDB code 2YOA) (main chain r.m.s.d. 0.43 Å, Extended Data Fig. 7b). We also established that the mutated SNARE complexes form properly (Extended Data Fig. 7f).

A subset of the interacting residues of region II were suggested to be functionally important in previous studies: Syt1 R398/R399 (refs 35, 36, 78) and SNAP-25 D51/E52/E55 (refs 27, 79, 80). However, interactions in region I have not been implicated in any previous studies.

CD spectroscopy of wild type and mutants of Syt1 C2B. CD measurements were conducted with CD spectrometer (Model 202-01, Aviv Biomedical, Inc.) equipped with a temperature controller. Data were collected with 10 µM samples of wild-type and mutant Syt1 C2B proteins in 20 mM Tris (pH 8.0), 150 mM NaCl buffer (without or with 5 mM CaCl₂) over a wavelength range of 195 nm to 260 nm, with 1 nm increments, in a 1 mm path length cell at 25 °C. Temperature denaturation experiments were performed at a wavelength of 216 nm by increasing the temperature from 25 °C to 100 °C in 3 °C temperature increments, a 2 min temperature equilibration time, and a 3 s averaging time. The fraction of unfolded protein at each temperature was calculated by using the formula $(I_{\text{obs}} - I_f)/(I_u - I_f)$, where I_{obs} is the observed mean residue ellipticity, and I_u and I_f are the mean residue ellipticities of the unfolded and folded states, respectively. I_u and I_f were estimated by extrapolation of the linear regions of the extremes of the denaturation curves.

Sequence alignment. The alignment was performed using ClustalW2 (<http://www.ebi.ac.uk/Tools/msa/clustalw2/>) and the figures were prepared with Boxshade3.21 (http://www.ch.embnet.org/software/BOX_form.html). List of UniProt or GenBank accession numbers of Syt1 homologues: *Caenorhabditis elegans* (worm Syt1, P34693); *Drosophila melanogaster* (fly Syt1, P21521); *Lymnaea stagnalis* (snail Syt1, AAO83847.1); *Doryteuthis pealeii* (squid Syt1, BAA09866.1); *Danio rerio* (zebrafish Syt1, XP_005164929.1); zebrafish Syt2, XP_009294914.1; zebrafish Syt9, AAI52175.1); *Xenopus tropicalis* (frog Syt1, XP_002935685.2); *Chelonia mydas* (turtle Syt1, XP_007057315.1); *Gallus gallus* (chicken Syt1, P47191); *Rattus norvegicus* (rat Syt1, P21707; rat Syt2, P29101; rat Syt9, P47861; rat Syt3, P40748; rat Syt4, P50232; rat Syt5, Q925C0; rat Syt6, Q62746; rat Syt7, Q62747; rat Syt8, Q925B4; rat Syt10, O08625; rat Syt11, O08835; rat Syt12, P97610; rat Syt13, Q925B5; rat Syt14, MOR7W7; rat Syt15, P59926; rat Syt16, D3ZB68); *Homo sapiens* (human Syt1, P21579; human Syt2, Q8N9I0; human Syt9, O00445). List of UniProt or GenBank accession numbers of SNAP-25 and syntaxin homologues: *Caenorhabditis elegans* (worm SNAP-25, NP_505641.2; worm syntaxin-1A, O16000); *Drosophila melanogaster* (fly SNAP-25, P36975; fly SNAP-29, NP_523831.1; fly syntaxin-1A, Q24547); *Lymnaea stagnalis* (snail syntaxin-1A, AAO83845.1); *Doryteuthis pealeii* (squid SNAP-25, AAM18191.1; squid syntaxin; CAA74913.1); *Danio rerio* (zebrafish SNAP-25, NP_001020729.1; zebrafish SNAP-29, NP_001243185.1; zebrafish syntaxin-1B, Q919P6); *Xenopus laevis* (frog SNAP-25, XP_005287463.1); *Xenopus tropicalis* (frog syntaxin-1A, NP_001072191.1); *Chrysemys picta bellii* (turtle SNAP-25, XP_007057315.1; turtle syntaxin-1A, XP_005294403.1); *Gallus gallus* (chicken SNAP-25, P60878; chicken syntaxin-1B, F5HN09); *Rattus norvegicus* (rat SNAP25a, P60881-2; rat SNAP-25b, P60881-1; rat SNAP-23, O70377; rat SNAP-29, Q9Z2P6; rat syntaxin-1A, P32851; rat syntaxin-1B, P61266; rat syntaxin-2, P50279-2; rat syntaxin-3, Q08849; rat syntaxin-4, Q08850; rat syntaxin-5, Q08851-2; rat syntaxin-7, O70257); *Homo sapiens* (human SNAP-25a, P60880-2; human SNAP-25b, P60880-1; human SNAP-23, O00161; human SNAP-29, O95721; human syntaxin-1A, Q16623-1); *Saccharomyces cerevisiae* (yeast sec-9, P40357; yeast sso-1, P32867).

Protein expression and purification for single vesicle-vesicle experiments. Full-length cysteine-free rat synaptobrevin-2, rat syntaxin-1A, rat SNAP-25A (with all endogenous cysteines changed to serines), rat Syt1 (with all endogenous cysteines changed to alanine, except the cysteine residue at position 277), and

wild-type rat complexin-1 were expressed and purified as previously described³⁸. All mutants of Syt1 and SNAP-25 were generated using the Quick Change Site-Directed Mutagenesis kit (Agilent) and expressed and purified using the same protocol.

Protein reconstitution for single vesicle-vesicle fusion experiments. Neuronal SNAREs and Syt1 represent a minimal system for Ca²⁺-triggered membrane fusion. Recent evidence for this notion came from a reconstituted single vesicle-vesicle assay that discriminates among vesicle association, lipid mixing and content mixing^{37,38,81}. Moreover, addition of complexin-1 greatly enhanced Ca²⁺-triggered amplitude and synchronization, and suppressed spontaneous release in this system³⁸. A variety of complexin-1 truncations and mutations qualitatively reproduced effects observed in neuronal cultures for both spontaneous and Ca²⁺-triggered release³⁸, lending credence to this reconstituted system to investigate mechanistic questions.

We used the same membrane compositions and protein densities as in our previous studies^{37,82}. Likewise, the reconstitution protocol was similar⁸² with several changes as described in ref. 38. Briefly, one class of vesicles was reconstituted with both Syt1, or its mutants, and synaptobrevin-2 to mimic synaptic vesicles (referred to as SV vesicles), while another class of vesicles was reconstituted with syntaxin-1A and SNAP-25 or its mutants to mimic plasma membranes (referred to as PM vesicles), using the previously described lipid compositions. The protein-to-lipid ratios used were 1:200 for synaptobrevin-2 and syntaxin-1A, and 1:1,000 for Syt1 and its mutants. A three- to fivefold excess of SNAP-25 and its mutants (with respect to syntaxin-1A) and 3.5 mol% PIP₂ were added to the protein-lipid mixture for PM vesicles only. Dried lipid films were dissolved in 110 mM β-octyl glucoside (β-OG) buffer containing purified proteins. Detergent-free buffer (20 mM HEPES-Na, pH 7.4, 90 mM NaCl, 0.1% 2-mercaptoethanol) was then added to the protein-lipid mixture until the β-OG concentration reached the critical micelle concentration 24.4 mM. The vesicles were then subjected to size-exclusion chromatography using a Sepharose CL-4B column, packed under near-constant pressure by gravity with a peristaltic pump (GE Healthcare) in a 5 ml column with a 2 ml bed volume, that was equilibrated with buffer V (20 mM HEPES-Na, pH 7.4, 90 mM NaCl, 20 µM EGTA, 0.1% 2-mercaptoethanol) followed by dialysis into 21 of detergent-free buffer V supplemented with 5 g of Bio-beads SM2 and 0.8 g l⁻¹ Chelex 100 resin (Bio-Rad, Life Science Research). After 4 h, the buffer was changed with 2 l of fresh buffer V containing Bio-beads and Chelex, and dialysis continued for 12 h. During the preparation of SV vesicles, 50 mM sulforhodamine B (Invitrogen) was present in all solutions before the size-exclusion chromatography step. As described previously⁸¹, the presence and purity of reconstituted proteins was confirmed by SDS-PAGE of the vesicle preparations, and the directionality of the membrane proteins (facing outward) was assessed by chymotrypsin digestion followed by SDS-PAGE. The size distributions of the SV and PM vesicles were analysed by cryo-EM, as described previously³⁷.

Single vesicle-vesicle content-mixing assay. We used the single vesicle-vesicle assay described in ref. 38. Briefly, SV vesicles were labelled with a soluble fluorescent content dye (sulforhodamine B) at a moderately self-quenching concentration; for simplicity in this work we did not include a lipid dye since we were exclusively interested in the exchange of content, the correlation for neurotransmitter release. The PM vesicles were immobilized on a surface that was passivated with polyethylene glycol (PEG) and functionalized via streptavidin-biotin linkages. SV vesicles were then added in the presence of 2 µM complexin-1. We directly started monitoring the arrival of SV vesicles to surface-immobilized PM vesicles during the first minute acquisition period. A stepwise increase in fluorescence emission of a spot in the field of view indicated the formation of a SV-PM vesicle pair during the vesicle association period.

Unbound SV vesicles were then removed through extensive washing with vesicle-free buffer, while continuing real-time observation of the fluorescence intensity; consequently we did not observe any additional SV-PM vesicle associations after the washing step. While continuing the observation for another 1-min period, a second step-wise increase of fluorescence intensity appeared for some fraction of the associated SV vesicles, which indicated Ca²⁺-independent, that is, spontaneous fusion events (referred to as the spontaneous fusion period). Next, we injected 500 µM Ca²⁺ solution, and continued monitoring for another minute, referred to as the Ca²⁺-triggered fusion period. For associated SV vesicles that did not undergo spontaneous fusion during the second period, a step-wise increase in fluorescence intensity during the third period indicated a Ca²⁺-triggered fusion event. To determine the temporal arrival of Ca²⁺ in the evanescent field of our TIR microscope setup, soluble Cy5 dye was added with the Ca²⁺ buffer to monitor the emergence of fluorescence intensity. Thus, our improved single vesicle-vesicle assay enables one to monitor the association of SV vesicles,

spontaneous, and Ca^{2+} triggered fusion events during the same data acquisition. Further details can be found in ref. 38.

Previous reconstitutions often employed lipid mixing, rather than content mixing, between vesicles with reconstituted neuronal SNAREs⁸³. For example, increased ensemble lipid mixing was observed upon addition of the soluble Syt1 C2AB fragment and Ca^{2+} (ref. 83). However, in retrospect, this result was probably caused by an effect on vesicle association since multivalent binding of the soluble Syt1 C2AB fragment can induce vesicle clustering¹⁹. Moreover, subsequent work revealed major differences between using the soluble C2AB fragment and reconstituted, full-length Syt1 (refs 84, 85). More importantly, assays based on lipid mixing alone can produce misleading results since lipid mixing can occur without content mixing⁸⁶. Our single vesicle-vesicle assay thus uses full-length Syt1 and it monitors content mixing, a correlate for neurotransmitter release^{37,38,81}.

Syt1 conditional knockout mice. Syt1 conditional knockout mice were generated by the European Conditional Mouse Mutagenesis Program (EUCOMM) and are available from the European Mouse Mutant Archive (EMMA) (EM:06829). Exon 2 of *Syt1* containing the transmembrane domain is floxed and its removal results in a frameshift that produces a truncated protein. The *Syt1* targeted mice were first crossed to FLPe mice to remove a gene trap cassette surrounded by *frt* sites. This cross yielded the conditional knockout mice as schematically shown in Fig. 3a. Exposure to Cre recombinase results in the total absence of synchronous release which is typical for Syt1 knockout neurons (Fig. 4a, b).

Neuronal cultures. Neuronal cultures were produced from wild-type and Syt1 conditional knockout mice as previously described⁸⁷. Hippocampi were dissected from P0 pups, dissociated by papain digestion, and plated on Matrigel-coated glass coverslips. Neurons were cultured *in vitro* in MEM supplemented with B27 (Gibco), glucose, transferrin, fetal bovine serum and Ara-C (Sigma), and were analysed after 14–16 days.

Lentivirus production. For rescue experiments, we used a lentiviral construct carrying a synapsin promoter, an optional rat *Syt1* cDNA, internal ribosome entry site (IRES), and a GFP-Cre recombinase fusion sequence. The control plasmid (TB592) contained no rescuing cDNA, with rescuing plasmids carrying the following cDNAs: TB761 (wild type), TB762 (R398Q/R399Q), TB765 (E295A/Y338W), TB767 (R281A/R398A/R399A), TB777 (R281A/E295A/Y338W/R398A/R399A). To make viruses, human embryonic kidney 293T cells were co-transfected with the lentiviral vector and three packaging plasmids. Supernatant containing the viruses was collected 48 h after transfection and was used to infect hippocampal neuronal cultures at day *in vitro* (DIV) four. Cultures were used for biochemical or physiological analyses at DIV 14–16.

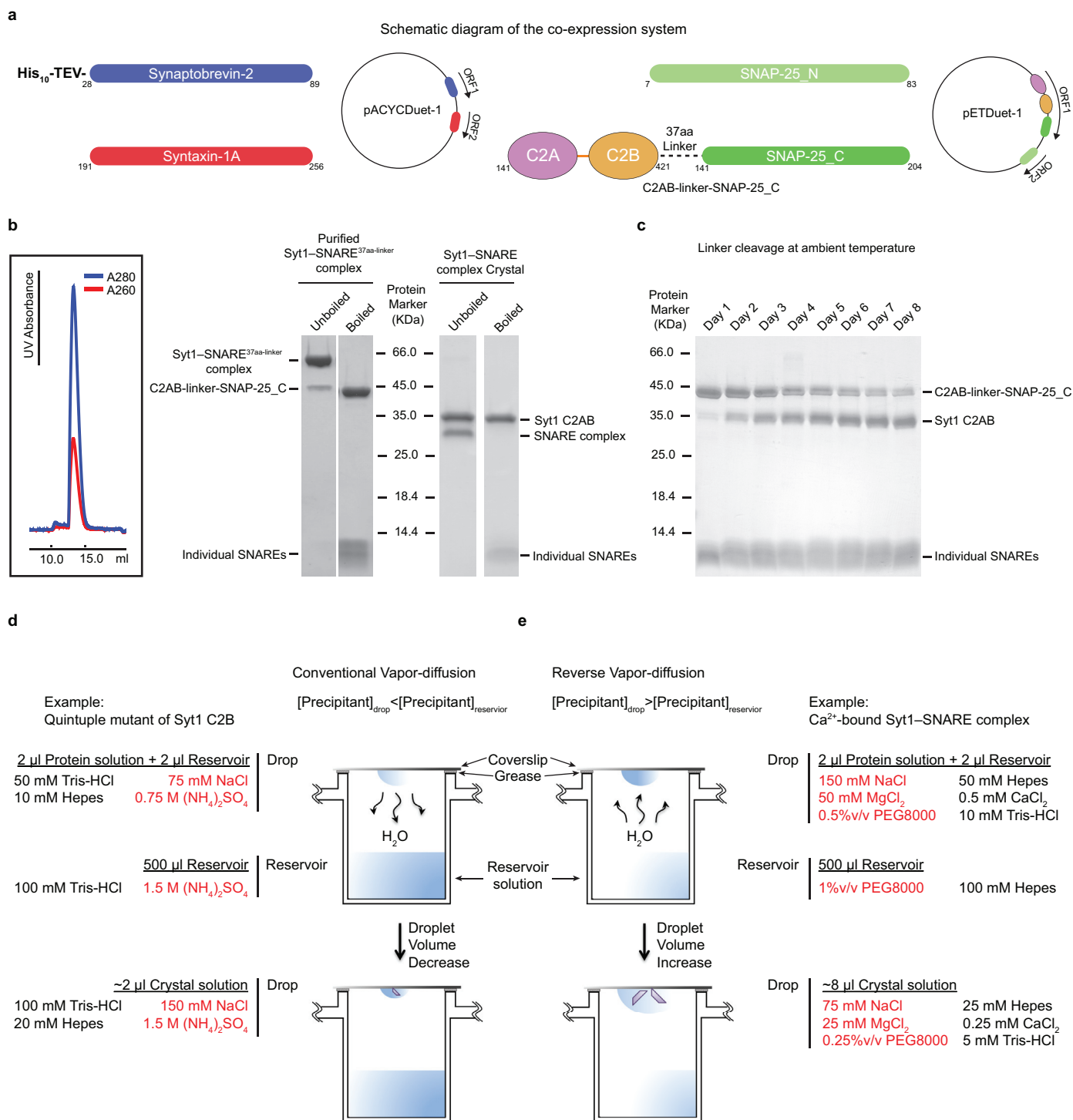
Electrophysiological recordings in cultured neurons. Recordings were performed essentially as previously described¹⁰. The whole-cell pipette solution contained 40 mM CsCl, 90 mM K-gluconate, 1.8 mM NaCl, 1.7 mM MgCl_2 , 3.5 mM KCl, 0.05 mM EGTA, 10 mM HEPES, 2 mM Mg-ATP, 0.4 mM Na-GTP, 10 mM phosphocreatine, and 10 mM QX-314 (pH 7.4, adjusted with CsOH). The bath solution contained 140 mM NaCl, 5 mM KCl, 2 mM CaCl_2 , 2 mM MgCl_2 , 10 mM HEPES, 10 mM glucose (pH 7.4, adjusted with NaOH). Evoked synaptic responses were triggered by a bipolar electrode. GABA-R-mediated IPSCs were pharmacologically isolated with CNQX (20 μM) and AP-5 (50 μM) in the bath solution and recorded at a -70 mV holding potential. Since the intracellular solution contains high internal Cl^- levels, IPSCs evoke large inward currents.

Immunoprecipitation and quantitative immunoblotting. Cultured Syt1 conditional knockout neurons infected with viruses expressing the desired mutants were solubilized in PBS (with 1 mM CaCl_2 , 0.2% Triton X-100, pH 7.4) supplemented EDTA-free protease inhibitor cocktail (Roche) for 1 h. The lysate was cleared by centrifugation at 16,000g for 10 min at 4 °C and immunoprecipitation was performed by incubating with polyclonal antibodies to syntaxin-1 (438B) or preimmune sera for 1 h at 4 °C, followed by incubation with 15 μl of a 50% slurry of protein-A Sepharose beads (GE Healthcare) for 2 h at 4 °C. Beads were washed 5 times with 1 ml extraction buffer, bound proteins were eluted with 2 \times SDS sample buffer containing 100 mM DTT and boiled for 20 min at 100 °C.

Co-precipitated proteins were separated by SDS-PAGE followed by detection with monoclonal antibodies against rat Syt1 (604.4, Synaptic Systems; this antibody does not detect mouse Syt1) and synaptobrevin-2 (cl. 69.1, Synaptic Systems). To allow for quantitative detection, dye-conjugated secondary antibodies were used (IRDye 800CW Donkey anti-Mouse IgG, Li-cor), membranes were scanned in an Odyssey scanner (Li-cor), and quantification was performed using Image Studio software (Li-cor). All experiments included a Syt1 wild-type group in addition to the desired mutants. The ratio of the Syt1/synaptobrevin-

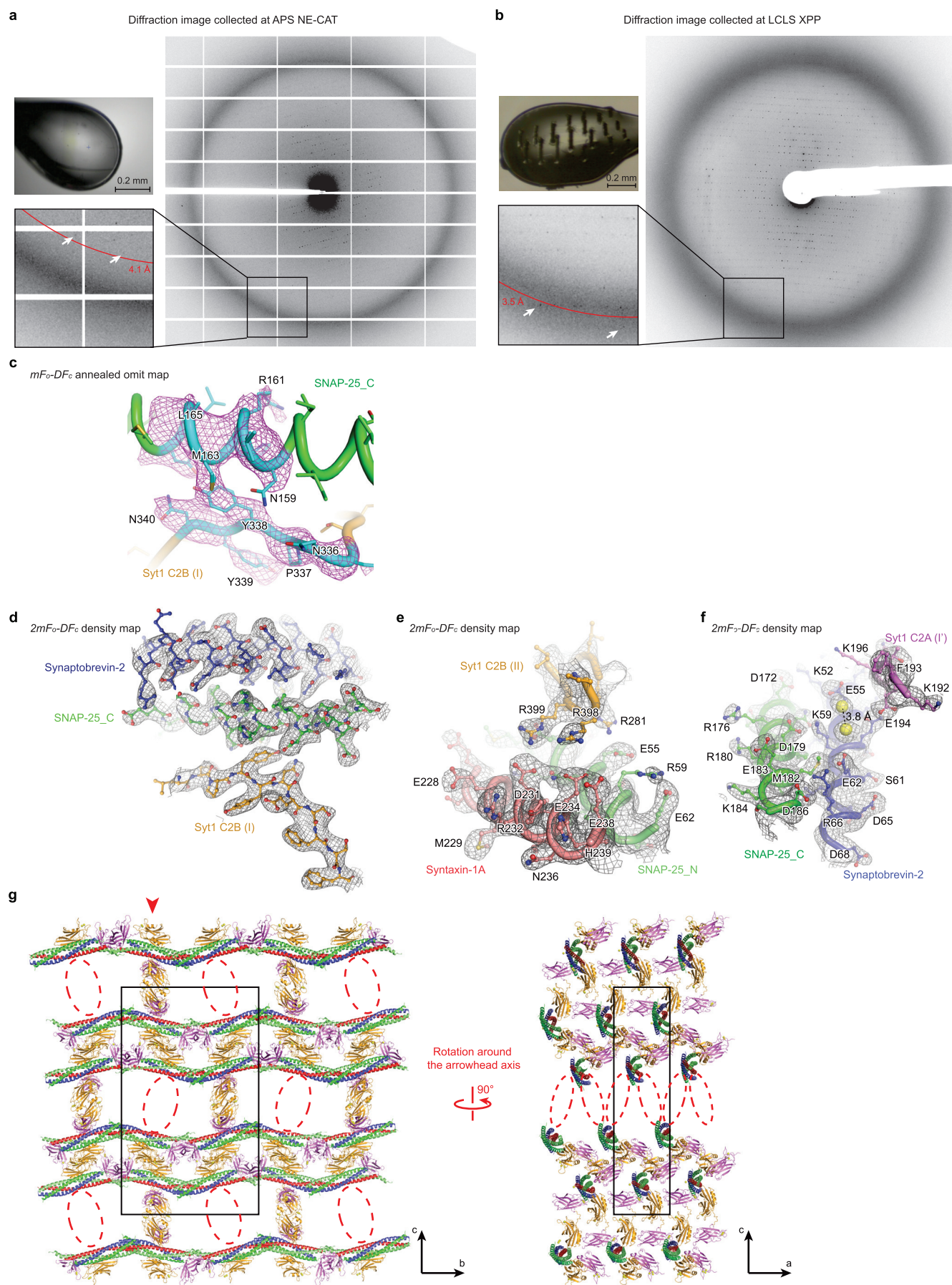
2 signal was determined and normalized with the Syt1 wild-type condition being set equal to 1.

51. van Leeuwen, H. C., Strating, M. J., Rensen, M., De Laat, W. & Van Der Vliet, P. C. Linker length and composition influence the flexibility of Oct-1 DNA binding. *EMBO J.* **16**, 2043–2053 (1997).
52. Ismail, S. A., Vetter, I. R., Sot, B. & Wittinghofer, A. The structure of an Arf-ArfGAP complex reveals a Ca^{2+} regulatory mechanism. *Cell* **141**, 812–821 (2010).
53. Ernst, J. A. & Brunger, A. T. High resolution structure, stability, and synaptotagmin binding of a truncated neuronal SNARE complex. *J. Biol. Chem.* **278**, 8630–8636 (2003).
54. Cipriano, D. J. et al. Processive ATP-driven disassembly of SNARE complexes by the N-ethylmaleimide sensitive factor molecular machine. *J. Biol. Chem.* (2013).
55. Studier, F. W. Protein production by auto-induction in high-density shaking cultures. *Protein Expr. Purif.* **41**, 207–234 (2005).
56. Bonifacio, R., Pellegrini, C. & Narducci, L. M. Collective instabilities and high-gain regime in a free electron laser. *Opt. Commun.* **50**, 373–378 (1984).
57. Kondratenko, A. M. & Saldin, E. L. Generation of coherent radiation by a relativistic electron beam in an undulator. *Part. Accel.* **10**, 207–216 (1980).
58. Neutze, R., Wouts, R., van der Spoel, D., Weckert, E. & Hajdu, J. Potential for biomolecular imaging with femtosecond X-ray pulses. *Nature* **406**, 752–757 (2000).
59. Solem, J. C. Imaging biological specimens with high-intensity soft x rays. *J. Opt. Soc. Am. B* **3**, 1551 (1986).
60. Cohen, A. E. et al. Goniometer-based femtosecond crystallography with X-ray free electron lasers. *Proc. Natl Acad. Sci. USA* **111**, 17122–17127 (2014).
61. Zeldin, O. B. et al. Data exploration toolkit for serial diffraction experiments. *Acta Crystallogr. D* **71**, 352–356 (2015).
62. Hattne, J. et al. Accurate macromolecular structures using minimal measurements from X-ray free-electron lasers. *Nature Methods* **11**, 545–548 (2014).
63. Sauter, N. K., Hattne, J., Grosse-Kunstleve, R. W. & Echols, N. New Python-based methods for data processing. *Acta Crystallogr. D* **69**, 1274–1282 (2013).
64. Sauter, N. K. et al. Improved crystal orientation and physical properties from single-shot XFEL stills. *Acta Crystallogr. D* **70**, 3299–3309 (2014).
65. Uervirojnangkoorn, M. et al. Enabling X-ray free electron laser crystallography for challenging biological systems from a limited number of crystals. *Elife* **4**, e05421 (2015).
66. Kabsch, W. XDS. *Acta Crystallogr. D* **66**, 125–132 (2010).
67. Winn, M. D. et al. Overview of the CCP4 suite and current developments. *Acta Crystallogr. D* **67**, 235–242 (2011).
68. McCoy, A. J. et al. Phaser crystallographic software. *J. Appl. Crystallogr.* **40**, 658–674 (2007).
69. Emsley, P. & Cowtan, K. Coot: Model-building tools for molecular graphics. *Acta Crystallogr. D* **60**, 2126–2132 (2004).
70. Brunger, A. T. Version 1.2 of the Crystallography and NMR system. *Nature Protocols* **2**, 2728–2733 (2007).
71. Brunger, A. T. et al. Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
72. Adams, P. D. et al. PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr. D* **58**, 1948–1954 (2002).
73. Schröder, G. F., Levitt, M. & Brunger, A. T. Super-resolution biomolecular crystallography with low-resolution data. *Nature* **464**, 1218–1222 (2010).
74. Urzhumtseva, L., Afonine, P. V., Adams, P. D. & Urzhumtsev, A. Crystallographic model quality at a glance. *Acta Crystallogr. D* **65**, 297–300 (2009).
75. Chen, V. B. et al. MolProbity: All-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66**, 12–21 (2010).
76. Pettersen, E. F. et al. UCSF Chimera - A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
77. Krissinel, E. & Henrick, K. Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* **372**, 774–797 (2007).
78. Young, S. M. & Neher, E. Synaptotagmin has an essential function in synaptic vesicle positioning for synchronous release in addition to its role as a calcium sensor. *Neuron* **63**, 482–496 (2009).
79. de Wit, H. et al. Synaptotagmin-1 docks secretory vesicles to syntaxin-1/SNAP-25 acceptor complexes. *Cell* **138**, 935–946 (2009).
80. Mohrmann, R. et al. Synaptotagmin interaction with SNAP-25 governs vesicle docking, priming, and fusion triggering. *J. Neurosci.* **33**, 14417–14430 (2013).
81. Kyoung, M. et al. *In vitro* system capable of differentiating fast Ca^{2+} -triggered content mixing from lipid exchange for mechanistic studies of neurotransmitter release. *Proc. Natl Acad. Sci. USA* **108**, E304–E313 (2011).
82. Kyoung, M., Zhang, Y., Diao, J., Chu, S. & Brunger, A. T. Studying calcium-triggered vesicle fusion in a single vesicle-vesicle content and lipid-mixing system. *Nature Protocols* **8**, 1–16 (2013).
83. Lee, H.-K. et al. Dynamic Ca^{2+} -dependent stimulation of vesicle fusion by membrane-anchored synaptotagmin 1. *Science* **328**, 760–763 (2010).
84. Lai, Y., Lou, X., Jho, Y., Yoon, T.-Y. & Shin, Y.-K. The synaptotagmin 1 linker may function as an electrostatic zipper that opens for docking but closes for fusion pore opening. *Biochem. J.* **456**, 25–33 (2013).
85. Wang, Z., Liu, H., Gu, Y. & Chapman, E. R. Reconstituted synaptotagmin I mediates vesicle docking, priming, and fusion. *J. Cell Biol.* **195**, 1159–1170 (2011).
86. Zick, M. & Wickner, W. T. A distinct tethering step is vital for vacuole membrane fusion. *Elife* **3**, e03251 (2014).
87. Maximov, A., Pang, Z. P., Tervo, D. G. R. & Südhof, T. C. Monitoring synaptic transmission in primary neuronal cultures using local extracellular stimulation. *J. Neurosci. Methods* **161**, 75–87 (2007).



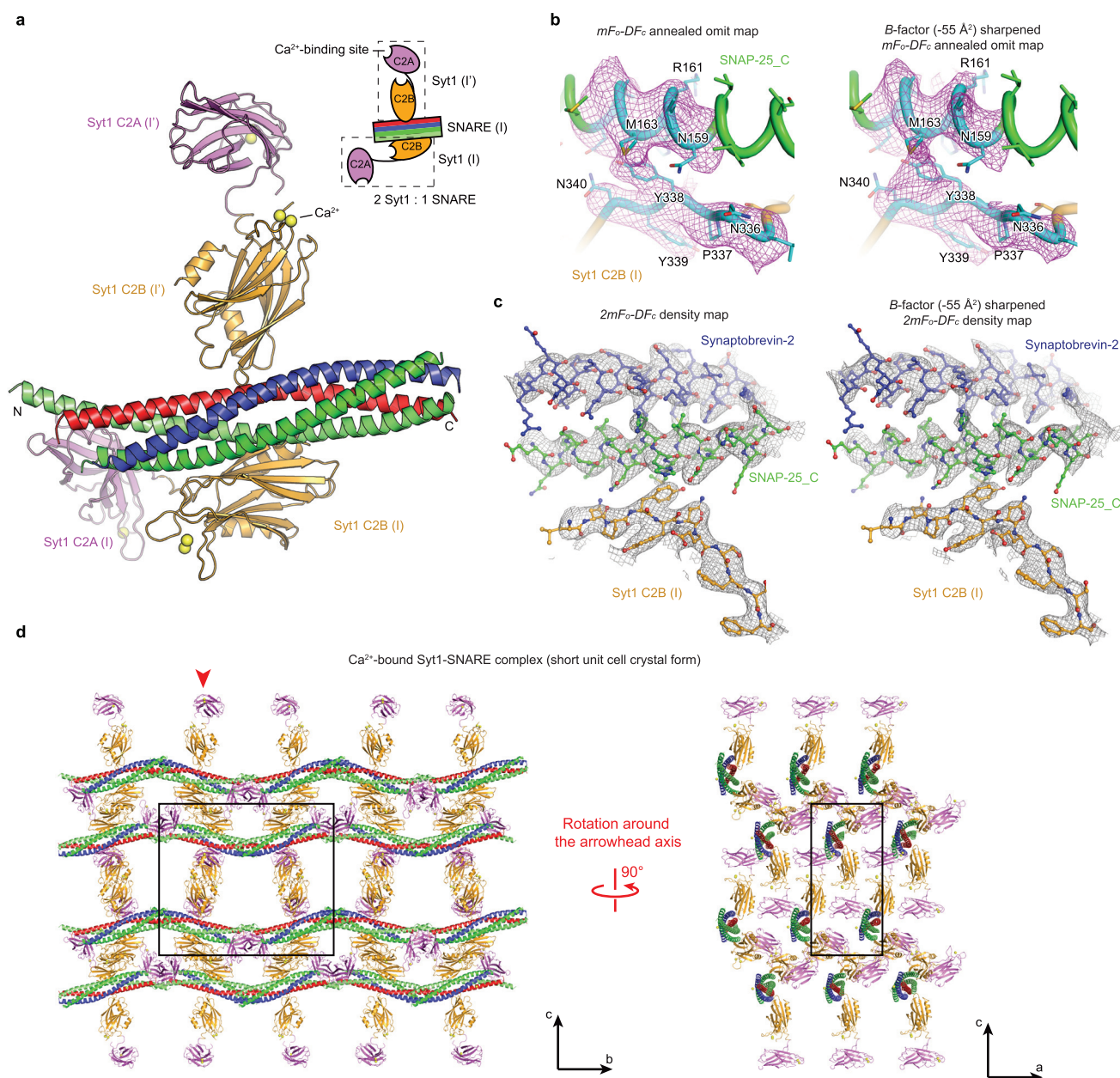
Extended Data Figure 1 | Purification and crystallization of the Syt1-SNARE complex. **a**, Diagram of the Duet co-expression vectors (Novagen) that express the fragments of the neuronal SNARE complex and the C2AB-linker-SNAP-25_C chimera used for purification and crystallization of the Syt1-SNARE^{37aa-linker} complex. The rat syntaxin-1A and His-tagged rat synaptobrevin-2 fragments were cloned into the vector pACYCDuet-1; the C2AB-linker-SNAP-25_C chimera and the SNAP-25_N fragment were cloned into the vector pETDuet-1 with amino acid ranges labelled. Dashed lines represent the 37-amino-acid linker (see Methods). **b**, The purified Syt1-SNARE^{37aa-linker} complex eluted as a single peak during size-exclusion chromatography (profile on the left). Left gel: Coomassie-blue-stained SDS-PAGE gel of the purified Syt1-SNARE^{37aa-linker} complex (unboiled and boiled). Right gel: Coomassie-blue-stained SDS-PAGE gel of dissolved crystals of the Syt1-SNARE complex that were grown over a period of 2 months starting

from purified Syt1-SNARE^{37aa-linker} (unboiled and boiled). Although Syt1 was initially covalently linked to SNAP-25_C, the linker was cleaved during crystallization. The comparison between boiled and unboiled lanes is a hallmark showing that neuronal SNARE complex is fully formed. **c**, Boiled Coomassie-blue-stained SDS-PAGE gel of the purified Syt1-SNARE^{37aa-linker} complex in solution at ambient temperature at the specified time after purification. Cleavage is apparent on day one and progresses slowly over several days. **d**, Schema showing the commonly used vapour-diffusion technique: the drop contains a lower concentration of the precipitant than the reservoir. The crystallization of the quintuple mutant of Syt1 C2B is used as an example. **e**, Schema showing a reverse vapour-diffusion method that was used for crystallization of the Ca^{2+} -bound Syt1-SNARE complex: the drop contains a higher concentration of the precipitant than the reservoir.



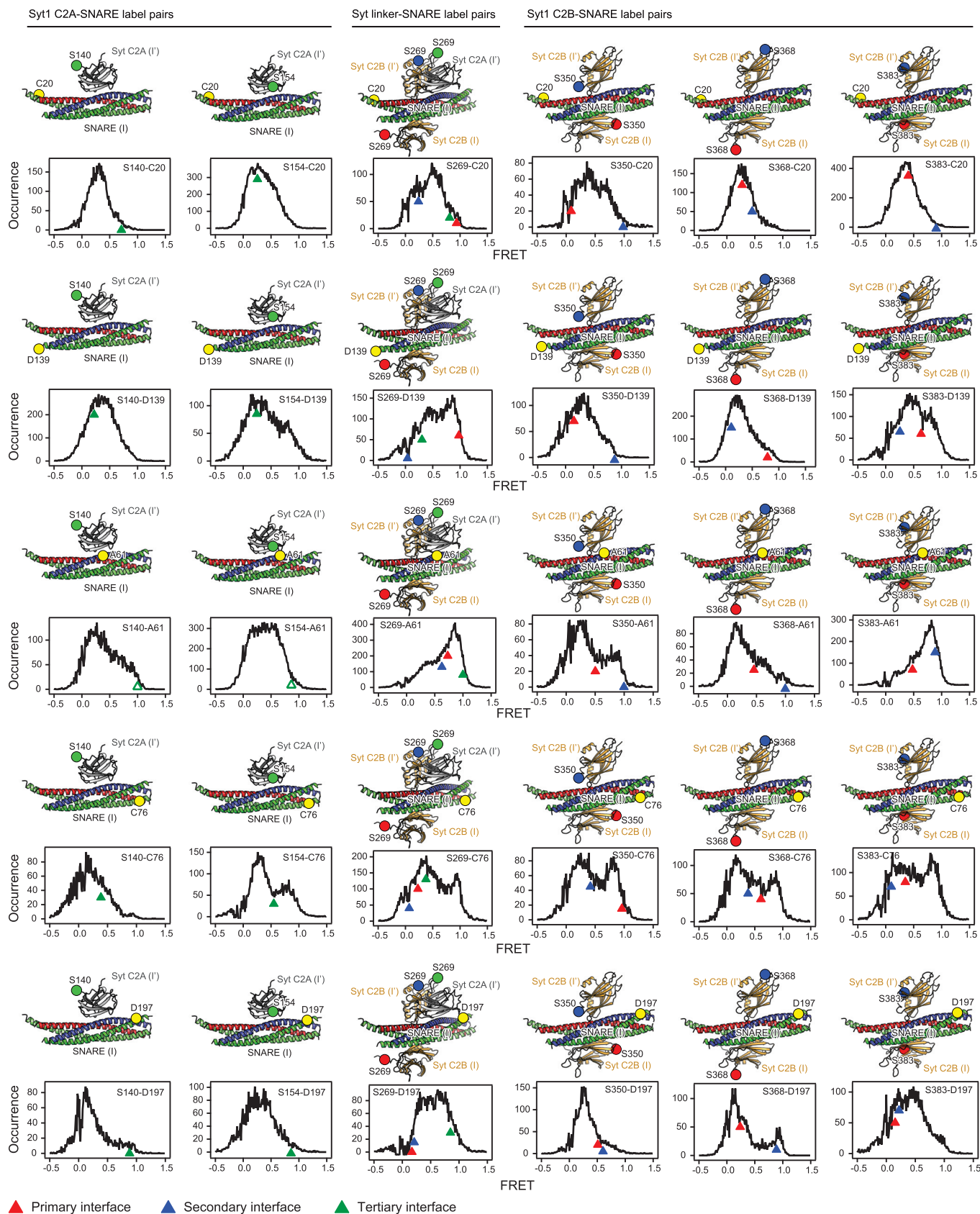
Extended Data Figure 2 | Diffraction images, electron density maps and crystal packing of the Syt1–SNARE complex in the long unit cell crystal form. **a**, Only one out of 85 screened crystals in the long unit cell crystal form diffracted to 4.1 Å resolution at the APS NE-CAT microfocus synchrotron beamline (a total of 105 crystals were screened with 20 that indexed in the short unit cell crystal form). **b**, A total of 61 out of ~72 crystals in the long unit cell crystal form diffracted to at least 3.5 Å resolution at the LCLS XFEL (a total of 148 crystals were diffracted, out of those 113 crystals produced 578 images that could be processed; 35 crystals did not diffract or showed multiple lattices). These exposures were taken along the crystal *c* axis. The left upper pictures in **a** and **b** show images of loop-mounted crystals after X-ray exposure. **c**, $mF_o - DF_c$ annealed omit map (Methods) of the Ca^{2+} -bound Syt1–SNARE complex in the long unit cell crystal form using diffraction data collected at

the LCLS XFEL; omitted residues within region I of the primary interface (residues 335–340 in Syt1 and 159–166 in SNAP-25) are coloured cyan. The contour level is 2.3σ . **d–f**, Representative $2mF_o - DF_c$ electron density maps of the Ca^{2+} -bound Syt1–SNARE complex in the long unit cell crystal form using diffraction data collected at the LCLS XFEL. The contour level is 1.5σ . **g**, Views of the crystal lattice perpendicular to the *bc* (left) and to the *ac* (right) planes of the Ca^{2+} -bound Syt1–SNARE complex in the long unit cell crystal form. The particular layer shown on the right corresponds to the red arrowhead in the left panel (only a slice corresponding to the layer is shown, creating the appearance of two disconnected groups of molecules—these groups are actually connected via interactions with the neighbouring layers). The red dashed oval indicates the ‘missing’ Syt1 C2AB fragment compared to the short unit cell crystal form (Extended Data Fig. 3d).



Extended Data Figure 3 | Asymmetric unit, electron density maps and crystal packing of the Syt1-SNARE complex in the short unit cell crystal form. **a**, Asymmetric unit of the Ca²⁺-bound Syt1-SNARE complex in the short unit cell crystal form at 3.6 Å resolution using diffraction data collected at the APS NE-CAT microfocus synchrotron beamline (Extended Data Table 1). The colour code is the same as in Fig. 1c. Two Syt1 C2AB fragments (distinguished by the designators I and I') bind to the same SNARE complex in the asymmetric unit (see schema). **b**, *mF_o - DF_c* annealed omit map (Methods) of the Ca²⁺-bound Syt1-SNARE complex in the short unit cell crystal form collected at the APS NE-CAT microfocus synchrotron beamline; omitted residues within region I of the primary interface (residues 335–340 in Syt1 and 159–166 in SNAP-25) are coloured cyan. The contour level is 2.3σ. Left side, without *B*-factor sharpening; right side, with *B*-factor sharpening. **c**, Representative *2mF_o - DF_c* electron density map of the Ca²⁺-bound

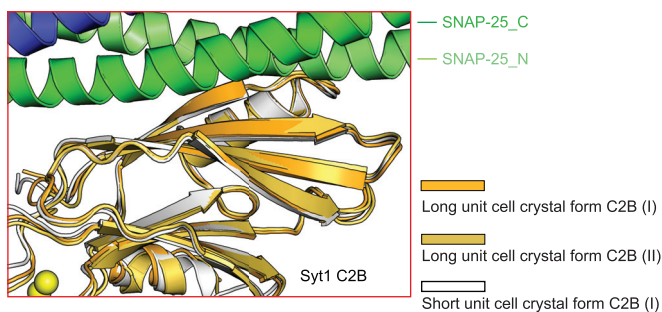
Syt1-SNARE complex for the short unit cell crystal form using diffraction data collected at the APS NE-CAT microfocus synchrotron beamline. The contour level is 1.5σ. Left side, without *B*-factor sharpening; right side, with *B*-factor sharpening. **b**, **c**, The sharpening *B*-factor (-55 \AA^2) was set to make the lowest atomic *B*-factor of the short unit cell crystal form comparable to that of the long unit cell crystal form. Even with *B*-factor sharpening, the electron density map of the long unit cell crystal form collected at the LCLS XFEL is superior to that of the short unit cell crystal form collected at the APS NE-CAT microfocus synchrotron beamline. **d**, Views of the crystal lattice perpendicular to the *bc* (left) and to the *ac* (right) planes of the Ca²⁺-bound Syt1-SNARE complex in the short unit cell crystal form. The particular layer shown on the right corresponds to the red arrowhead in the left panel. The unit cell is outlined by a black box.



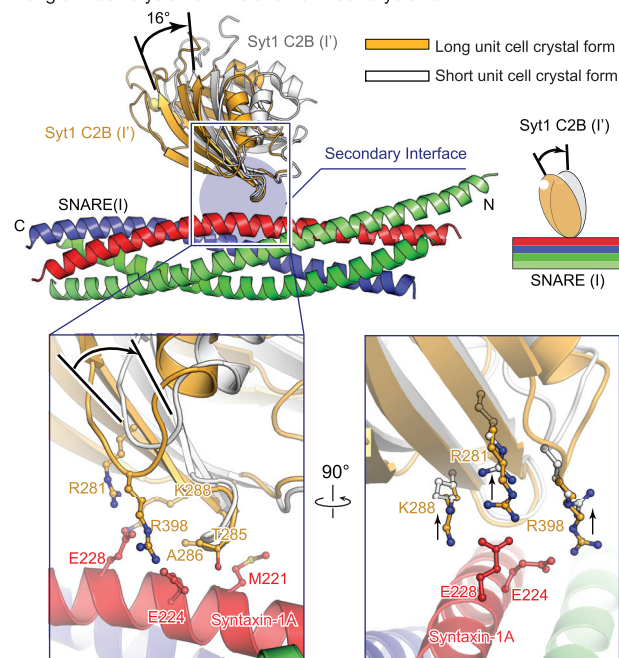
Extended Data Figure 4 | Single-molecule FRET efficiency distributions of the Syt1–SNARE complex versus FRET efficiency values calculated from the Syt1–SNARE interfaces observed in the crystal structure. Shown are histograms of intermolecular single molecule FRET (smFRET) efficiency values that were measured between pairs of covalently attached organic labels on the Syt1 C2AB fragment and the SNARE complex²⁸ (also shown as large spheres superimposed on the interfaces observed in the crystal structure). Arrowheads indicate FRET efficiencies calculated from the crystal structure of the Ca^{2+} -bound Syt1–SNARE complex in the long unit cell crystal form (complex I) for the primary, secondary and tertiary interfaces, using the methods and approximations described in ref. 28 to simulate the positions of dye centres in order to calculate the FRET-efficiency values. Only the dye pair

combinations between the nearest C2 domain (including the C2A–C2B linker) and the SNARE complex were calculated for the three interfaces. Note that owing to the presence of transitions between different states the histograms reflect a combined effect of interaction interfaces. The label at position A61 would have disrupted the tertiary interfaces between the C2A domain and the SNARE complex, explaining the discrepancy for these labels (indicated by open triangles). In retrospect, the top smFRET-derived model²⁸ and the primary interface observed in the crystal structure primarily differed in the orientation of the C2B domain. Moreover, the top smFRET-derived model predicted the approximate location of the primary interface on the neuronal SNARE (see Fig. 4c in ref. 28).

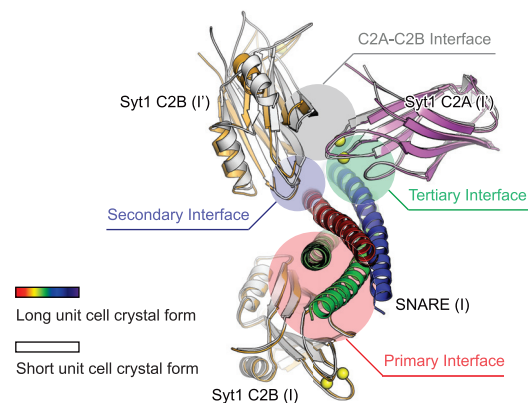
a Primary interface
long unit cell crystal form vs short unit cell crystal form



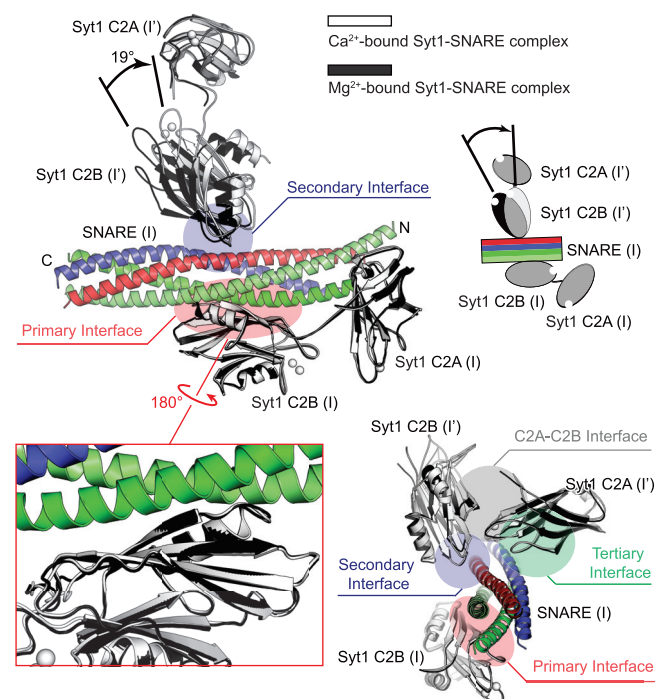
b Secondary interface
long unit cell crystal form vs short unit cell crystal form



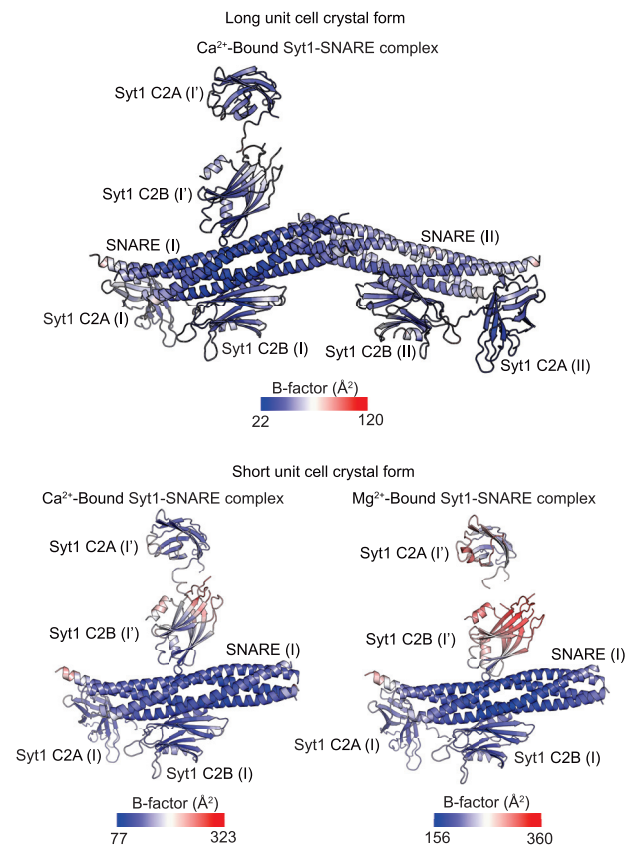
c Other interfaces
long unit cell crystal form vs short unit cell crystal form



d Ca²⁺-bound vs Mg²⁺-bound Syt1-SNARE complex
(Short unit cell crystal form)



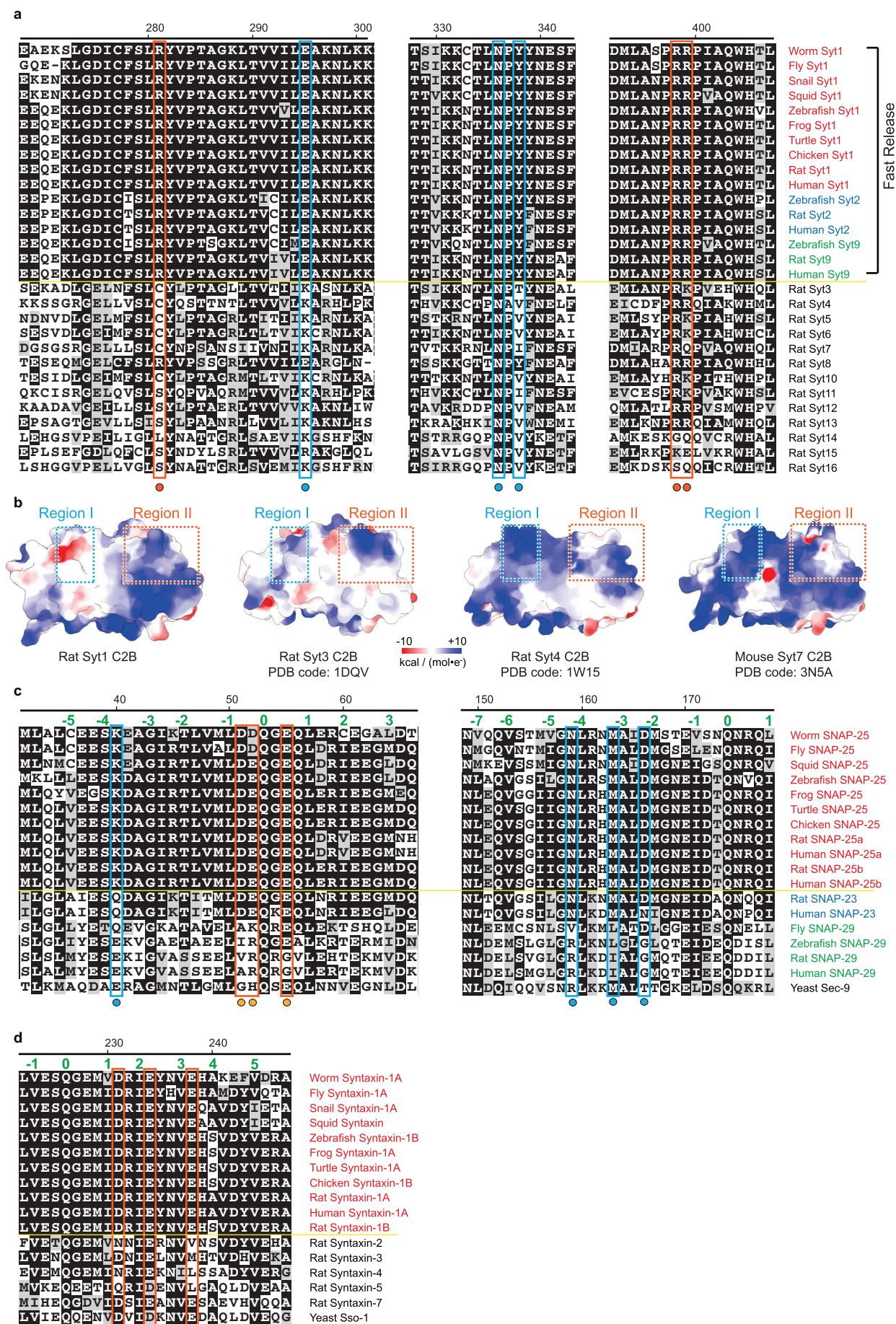
e



Extended Data Figure 5 | Comparison of the two crystal forms and the Ca^{2+} - and Mg^{2+} -bound crystal structures of the Syt1–SNARE complex.

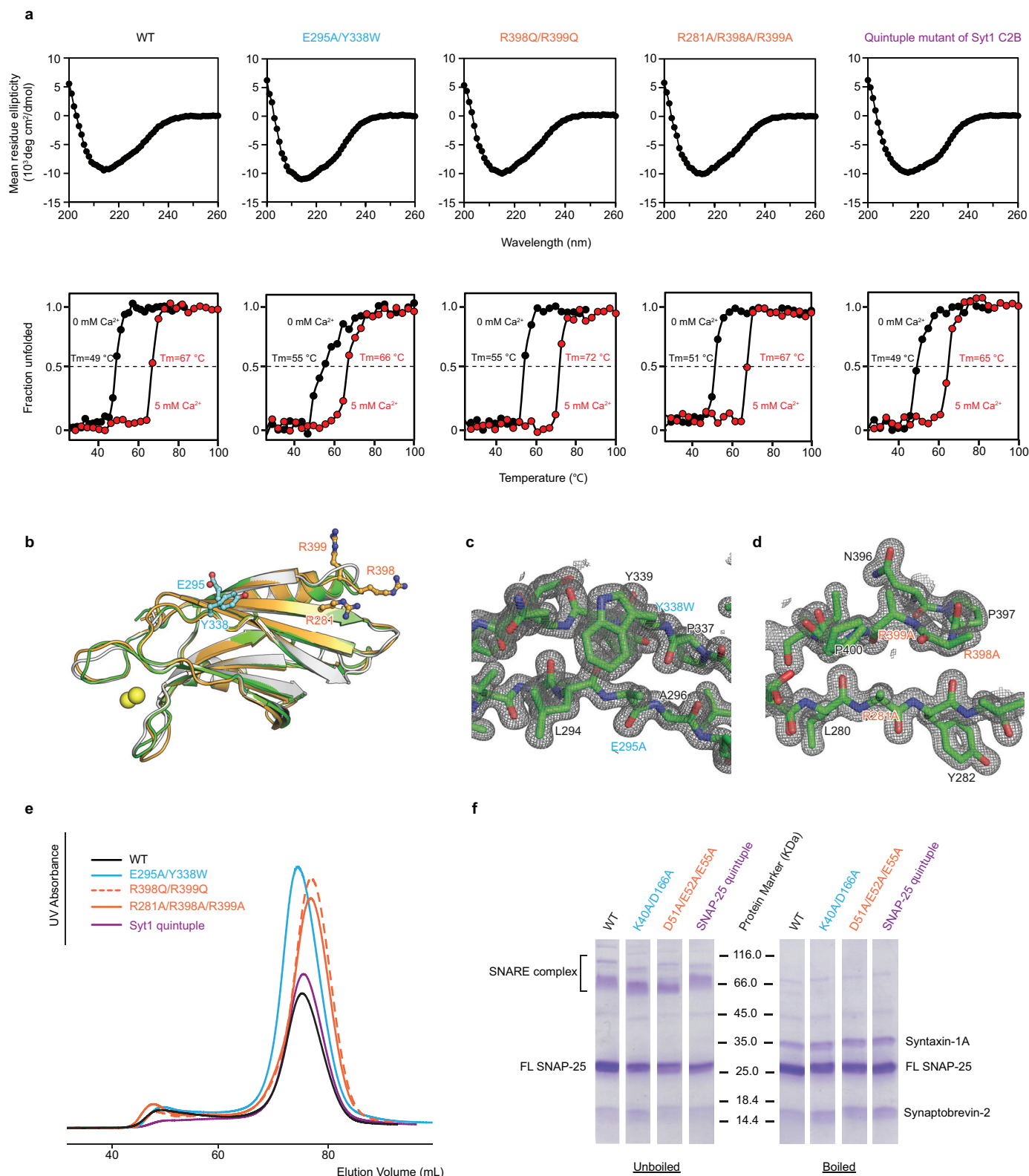
a, Superposition of the primary interfaces of the Ca^{2+} -bound Syt1–SNARE complex structure in the long unit cell crystal form (gold and bright orange) and in the short unit cell crystal form (white). The primary interface is very similar in both crystal forms: the r.m.s.d. for the primary interface between both crystal forms is 0.38 Å (bright orange) and 0.42 Å (white) for complex I and complex II, respectively (including $\text{C}\alpha$ atoms of the SNARE complex and the Syt1 C2B (I) domain forming the interface). **b**, Superposition of complex I in the long unit cell crystal form with the asymmetric unit of the short unit cell crystal form, but only showing the secondary interface (light-blue shaded disk) between Syt1 C2B (I') and the SNARE complex (I). The bottom panels show close-up views of the secondary interface: left, interacting residues (sticks and balls); right, a 90° rotated view of the view shown in the left panel. The Syt1 C2B (I') domain is rotated by 16° between the two crystal forms and, as a consequence, the interactions between residues R281, K288 and R398 of the Syt1 C2B (I') domain and residues E224 and E228 of syntaxin-1A are slightly changed by this rotation. Notably, residues Syt1 R281, K288 and R398 are involved in both the primary (Fig. 2) and secondary interfaces.

c, Superposition of complex I in the long unit cell crystal form with the asymmetric unit of the short unit cell crystal form, showing all interfaces. **d**, Superposition of the Ca^{2+} -bound (white) and Mg^{2+} -bound (black) crystal structures of the Syt1–SNARE complex, both in the short unit cell crystal form. The lower left panel shows a close-up view of the primary interface, indicating that it is very similar in both the Ca^{2+} - and Mg^{2+} -bound crystal structures. The Syt1 C2B domain that forms the secondary interface (light-blue shaded disk) is rotated by 19° between the Ca^{2+} - and Mg^{2+} -bound complexes. The lower-right panel is a rotated view of the complex, also showing the tertiary interface (light-green shaded disk), and the C2A–C2B interface that involves asymmetry-related Syt1 C2A domain (I') (grey shaded disk). **e**, *B*-factor coloured cartoon representations of the asymmetric units of the Ca^{2+} -bound long unit cell crystal form (top), the Ca^{2+} -bound short unit cell crystal form (bottom left), and the Mg^{2+} -bound short unit cell crystal form (bottom right) of the Syt1–SNARE complex. Note that the primary interfaces have relatively low *B*-factors, similar to the majority of the structure, while parts of the C2A and C2B domains involved in the secondary and tertiary interfaces have higher *B*-factors, possibly indicating increased flexibility.



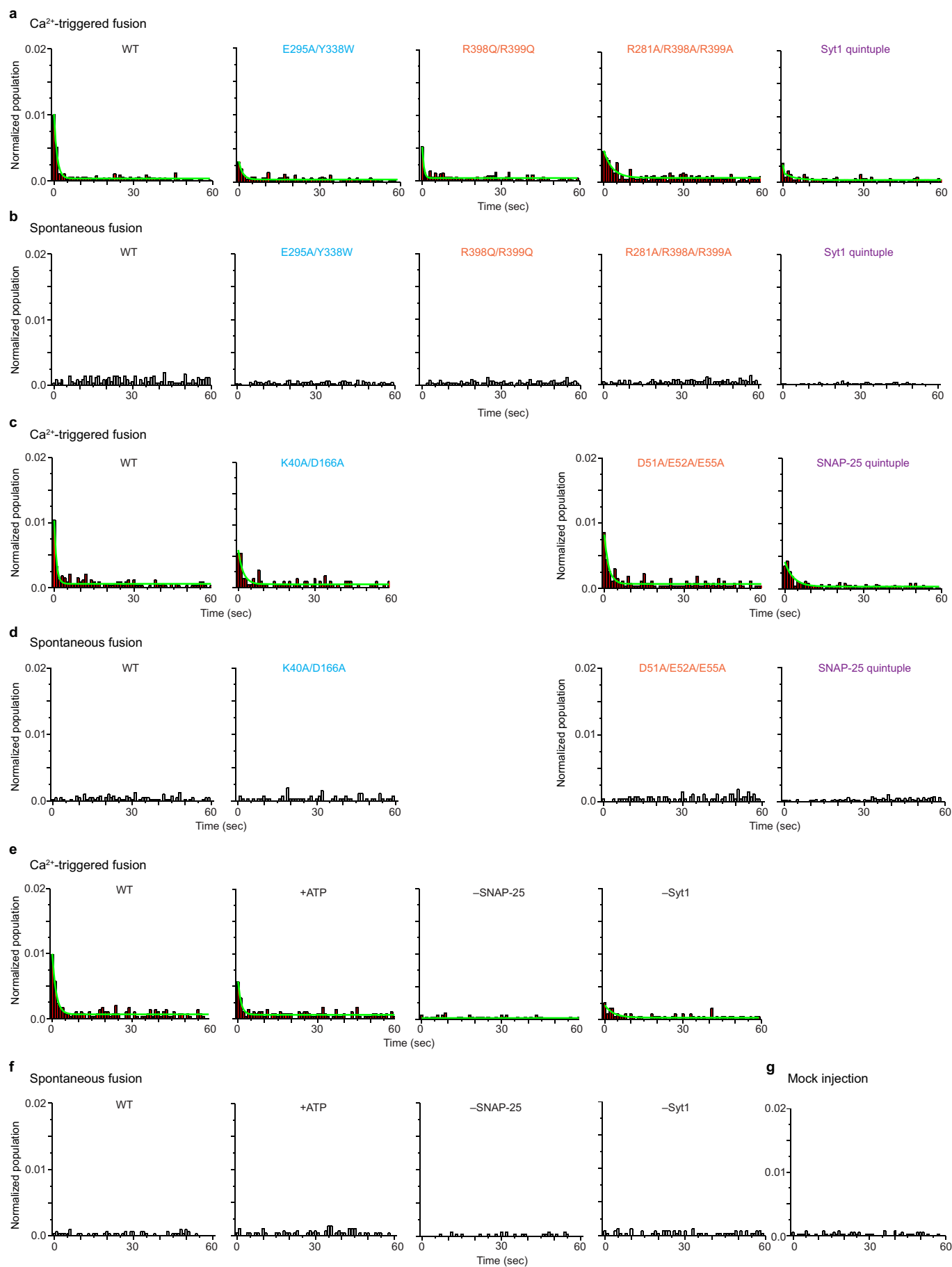
Extended Data Figure 6 | Sequence alignments of Syt1, SNAP-25 and syntaxin-1A from different homologues. **a**, Sequence alignment of Syt1 homologues, showing the sequences around the primary interface of the Syt1–SNARE complex. Note that rat Syt5 refers to UniProt ID Q925C0, zebrafish Syt9 refers to GeneBank accession number AAI52175, rat Syt9 refers to UniProt ID P47861, and human Syt9 refers to UniProt ID O00445. **b**, Electrostatic potential surfaces of the known crystal structures of synaptotagmin-1, synaptotagmin-3, synaptotagmin-4 and synaptotagmin-7;

the dashed rectangles indicate the regions that correspond to the primary interface regions I and II of the Syt1–SNARE complex. **c**, Sequence alignment of different SNAP-25 homologues, showing the sequences around the primary interface of the Syt1–SNARE complex. **d**, Sequence alignment of different syntaxin homologues, showing a sequence range around the primary interface of the Syt1–SNARE complex. In all panels, the interacting residues of the primary interface are indicated by solid circles and coloured boxes for region I (cyan) and region II (red/orange).



Extended Data Figure 7 | Syt1 mutants and SNARE complexes with SNAP-25 mutants are well folded. **a**, Top panels: CD spectra of wild-type and mutant Syt1 C2B domains in the absence of Ca^{2+} . Bottom panels: thermal denaturation was monitored by molar ellipticity at a wavelength of 216 nm in the absence of Ca^{2+} (black) and in the presence of 5 mM Ca^{2+} (red). The specified melting temperatures were estimated as the mid-point of the melting curves (Methods). **b**, Superposition of the Syt1 C2B domains from the Ca^{2+} -bound Syt1-SNARE complex in the short unit cell crystal form (gold), the crystal structure of the quintuple mutant (R281A/E295A/Y338W/R398A/R399A) of the Syt1 C2B domain (green), and the crystal structure of the isolated

Syt1 C2B domain (white, PDB code 2YOA). **c**, **d**, Representative $m2F_o - DF_c$ electron density maps of the crystal structure of the quintuple mutant of the Syt1 C2B domain (Extended Data Table 1) contoured at 2.0σ . The labels refer to the mutated residues. **e**, Overlay of SEC profiles of full-length Syt1 mutant proteins used in the single vesicle-vesicle fusion assay (Fig. 3d-g). **f**, Coomassie-blue-stained SDS-PAGE with and without boiling of neuronal SNARE complexes formed by full-length SNAP-25 and its mutants, syntaxin-1A and synaptobrevin-2, using the proteins that were used in the single vesicle-vesicle fusion assay (Methods).



Extended Data Figure 8 | Probability of fusion versus time upon 500 μM Ca^{2+} injection and spontaneous fusion for Syt1 and SNAP-25 mutants. Shown are the data that were used to generate Fig. 3d–g. The number of independent experiments and analysed events are provided in Extended Data Table 2. **a–d**, Cumulative histograms of probability of fusion versus time

for Syt1 mutants upon 500 μM Ca^{2+} injection (**a**) and spontaneous fusion (**b**), and SNAP-25 mutants upon 500 μM Ca^{2+} injection (**c**) and spontaneous fusion (**d**). **e–g**, Control experiments: **e**, Ca^{2+} -triggered fusion; **f**, spontaneous fusion with 3 mM ATP, without SNAP-25 or Syt1; and **g**, mock injection without Ca^{2+} .

Extended Data Table 1 | Crystallographic data and refinement statistics

	Ca ²⁺ -bound Syt1– SNARE complex (long unit cell crystal form)	Ca ²⁺ -bound Syt1– SNARE complex (short unit cell crystal form)	Mg ²⁺ -bound Syt1– SNARE complex (short unit cell crystal form)	Quintuple mutant of Syt1 C2B
Data collection				
Beamline	SLAC-LCLS	APS-NECAT	APS-NECAT	APS-NECAT
Space group	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2	P2 ₁ 2 ₁ 2	P2 ₁ 2 ₁ 2 ₁
Cell dimensions				
<i>a</i> , <i>b</i> , <i>c</i> (Å)	69.6, 171.1, 291.9	69.1, 171.6, 146.9	69.1, 171.8, 146.6	68.9, 107.6, 110.8
α , β , γ (°)	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0
Resolution (Å)	20.0–3.50 (3.62–3.50)*	50.0–3.60 (3.73–3.60)	85.9–4.10 (4.32–4.10)	50.0–1.65 (1.69–1.65)
<i>R</i> _{merge} (%) (rotation) [†]	--	6.7 (74.1)	10.4 (64.6)	6.40 (91.7)
<i>R</i> _{merge} (%) (still) [†]	39.7 (32.2)	--	--	--
<i>CC</i> _{1/2}	92.7 (35.5)	99.9 (89.5)	99.7 (64.6)	99.9 (78.3)
<i>I</i> / σ <i>I</i>	7.7 (2.6)	12.3 (1.9)	8.8 (1.9)	18 (2.3)
Completeness (%)	87.6 (65.6)	99.7 (98.7)	95.9 (97.0)	99.0 (98.5)
Multiplicity (rotation) [†]	--	14.5 (12.4)	3.4 (3.5)	6.7 (6.8)
Multiplicity (still) [†]	5.0 (1.8)	--	--	--
Refinement				
Resolution (Å)	20.0–3.50 (3.62–3.50)	50.0–3.60 (3.73–3.60)	50.0–4.10 (4.25–4.10)	50.0–1.65 (1.71–1.65)
No. reflections	39174 (2884)	20846 (2004)	13519 (1320)	98575 (9676)
<i>R</i> _{work} / <i>R</i> _{free}	0.322 / 0.353	0.249 / 0.289	0.276 / 0.323	0.150 / 0.178
No. atoms				
Protein	10890	6506	6510	5033
Ca ²⁺	19	7	0	0
Mg ²⁺	0	0	4	0
<i>B</i>-factors				
Protein	49	158	194	23.1
Ca ²⁺	32	187	--	--
Mg ²⁺	--	--	202	--
R.m.s. deviations				
Bond lengths (Å)	0.003	0.004	0.003	0.019
Bond angles (°)	0.758	0.862	0.714	1.763

*Values in parentheses are for the highest resolution shell.

[†]“(rotation)” refers to rotation diffraction data collected at the APS synchrotron and “(still)” refers to still diffraction data collected at the LCLS XFEL.

Extended Data Table 2 | Data summary table for the single vesicle-vesicle fusion experiments with Syt1 and SNAP-25 mutants

	No. of spontaneous fusion events	No. of Ca ²⁺ triggered fusion events	Total no. of analyzed events (no. of associated vesicle pairs)	No. of independent experiments (N)
Syt1 WT	90	167	3764	5
	100	262	4632	6
	49	152	3035	5
	64	173	3564	4
	159	256	6503	9
<i>Syt1 mutants</i>				
E295A/Y338W	113	246	5872	12
R398Q/R399Q	166	217	6724	7
R281A/R398A/R399A	133	271	5134	7
Syt1 quintuple	49	143	5241	7
<i>SNAP-25 mutants</i>				
K40A/D166A	61	139	2549	7
D51A/E52A/E55A	81	150	2717	7
SNAP-25 quintuple	105	261	6741	9
<i>Control</i>				
+ATP	59	129	2830	5
−SNAP-25	24	36	3040	5
−Syt1	57	58	2379	5
Mock	--	25	1947	5

Structural insights into the bacterial carbon–phosphorus lyase machinery

Paulina Seweryn¹, Lan Bich Van¹, Morten Kjeldgaard¹, Christopher J. Russo², Lori A. Passmore², Bjarne Hove-Jensen¹, Bjarne Jochimsen¹ & Ditlev E. Brodersen¹

Phosphorus is required for all life and microorganisms can extract it from their environment through several metabolic pathways. When phosphate is in limited supply, some bacteria are able to use phosphonate compounds, which require specialized enzymatic machinery to break the stable carbon–phosphorus (C–P) bond. Despite its importance, the details of how this machinery catabolizes phosphonates remain unknown. Here we determine the crystal structure of the 240-kilodalton *Escherichia coli* C–P lyase core complex (PhnG–PhnH–PhnI–PhnJ; PhnGHIJ), and show that it is a two-fold symmetric hetero-octamer comprising an intertwined network of subunits with unexpected self-homologies. It contains two potential active sites that probably couple phosphonate compounds to ATP and subsequently hydrolyse the C–P bond. We map the binding site of PhnK on the complex using electron microscopy, and show that it binds to a conserved insertion domain of PhnJ. Our results provide a structural basis for understanding microbial phosphonate breakdown.

Phosphonate compounds that contain a stable C–P bond are used as a source of phosphate by microorganisms in many natural environments where low levels of free and organic phosphate limit growth¹. The C–P lyase pathway, which converts phosphonate into 5-phosphoribosyl- α -1-diphosphate (PRPP) in an ATP-dependent fashion, is activated upon phosphate starvation in many bacterial species including *Escherichia coli*^{2,3}. The enzymes of this pathway have a very broad substrate specificity enabling the bacteria to utilize a wide range of compounds for growth including alkyl, amino-alkyl and aryl phosphonates^{4–7}.

In *E. coli*, the 14-cistron *phn* operon is required for phosphonate uptake and utilization and encodes an ATP-binding cassette transporter (PhnC, PhnD and PhnE), a regulatory protein (PhnF) and components required for enzymatic conversion of phosphonate into PRPP (PhnGHJKLMN^{8–13}). PhnG, PhnH, PhnI and PhnJ have been shown to form a stable protein complex, which we term the C–P lyase core complex, probably with PhnG and PhnI at its centre^{14,15}. The core complex stably associates with a fifth protein, PhnK, which resembles ABC cassette proteins, with unknown stoichiometry¹⁴. PhnJ contains an iron–sulfur cluster required for C–P bond cleavage via an *S*-adenosyl methionine (SAM)-dependent radical mechanism^{16–18}, while PhnI is a nucleosidase capable of deglycosylating ATP and GTP to ribose 5-triphosphate¹⁸. A reaction mechanism for the breakdown of phosphonate via the C–P lyase pathway was proposed where PhnI, supported by PhnG, PhnH and PhnL, the latter a protein not present in the core complex, catalyses the transfer of the phosphonate moiety to the ribose 1' position of ATP by displacing adenine, generating a ribose 5'-triphosphate alkyl phosphonate intermediate (Fig. 1a and Extended Data Fig. 1). Following pyrophosphate release by PhnM, PhnJ cleaves the C–P bond and PhnP/PhnN convert the resulting ribose cyclic phosphate into PRPP^{10,11}. The C–P lyase core complex thus harbours two key activities of this pathway; coupling of the phosphonate to ATP (PhnG, PhnH and PhnI) and C–P bond cleavage (PhnJ)¹⁸. PhnH is the only component of the C–P lyase core complex that has been structurally characterized and displays a fold related to the pyridoxal-5'-phosphate-dependent transferases.

It forms a homodimer when expressed independently¹⁹, and its role within the complex is unclear.

The global architecture of C–P lyase

We purified the *E. coli* C–P lyase core complex and determined its crystal structure by molecular replacement in combination with single-wavelength anomalous dispersion using a Ta₆Br₁₂ cluster derivative and the PhnH crystal structure¹⁹ as a search model. The structure was refined using a native data set extending to 1.7 Å with resulting final *R* factors of 14.9% (*R*_{work}) and 17.6% (*R*_{free}) (Extended Data Table 1 and Extended Data Fig. 2). The structure consists of two copies of each of PhnG (16 kDa), PhnH (21 kDa), PhnI (39 kDa) and PhnJ (32 kDa), comprising a total of 1,958 amino acid residues in the asymmetric unit (Fig. 1b and Extended Data Fig. 3), and is complete except for a few residues located at the subunit termini. The structure includes four sulfate ions, four zinc ions and 1,792 solvent molecules. Together, the eight polypeptides form a compact and intertwined, two-fold symmetric hetero-octamer that can be described as (PhnGHIJ)₂, with a total molecular mass of 240 kDa (Fig. 1c, d), consistent with its behaviour in solution¹⁴.

The C–P lyase core complex resembles the letter 'H' with rounded arms that are twisted approximately 45° in and out of the plane with respect to each other. The arms are composed on opposing sides by the two PhnG molecules and on the other sides by tight complexes between PhnJ and PhnH (Fig. 1c). At the centre of the molecule, a compact PhnI homodimer forms a disc-like structure that serves as a central hub for attachment of the other subunits (Fig. 1c, d, green). The core domain of PhnI, which is the largest single domain in the structure, has a novel $\alpha + \beta$ fold comprised of a four-stranded, anti-parallel β -sheet next to a four-helix bundle combining in a unique fold: the β -barrel domain (Figs 1c and 2a). At both termini there are helical extensions of approximately 35 residues that grasp PhnJ and tether it to the complex via extensive interactions (Fig. 3a). In turn, PhnJ attaches PhnH to the complex through packing of conserved α -helices in both proteins (Fig. 1c).

¹Department of Molecular Biology and Genetics, Aarhus University, Gustav Wieds Vej 10c, DK-8000 Aarhus C, Denmark. ²Medical Research Council Laboratory of Molecular Biology, Francis Crick Avenue, Cambridge CB2 0QH, UK.

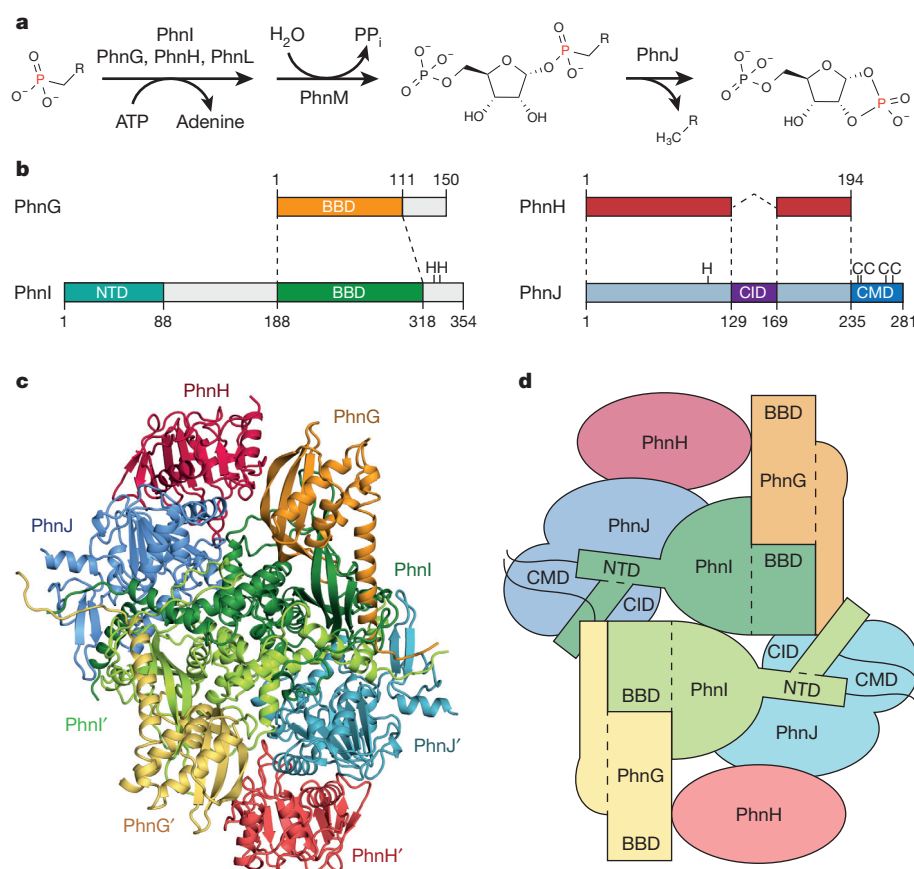


Figure 1 | Overall architecture of the C-P lyase core complex. **a**, The C-P lyase core complex catalyses transfer of a phosphonate to the 1' position of ATP (PhnI assisted by PhnG, PhnH and PhnL) and cleavage of the C-P bond (PhnJ). **b**, Overview of the four proteins with dashed lines indicating conserved structural domains. Functional residues are shown with amino acid one-letter code. BBD, β -barrel domain; CID, central insertion domain; CMD, C-terminal mini domain; NTD, N-terminal domain. **c**, Overall structure of the 240 kDa C-P lyase core complex. **d**, Schematic architecture of the complex with structural domains indicated.

PhnJ has a compact $\alpha + \beta$ fold surrounded by two mini-domains, the central insertion domain (CID) and the C-terminal mini domain (CMD) (Figs 1b, 2b and Extended Data Fig. 3). Surprisingly, the core folds of PhnJ and PhnH are nearly identical ($C\alpha$ r.m.s.d. 2.5 Å), despite very little sequence similarity (Extended Data Fig. 4a). Moreover, the interactions in the PhnH-PhnJ heterodimer closely resemble those observed in the crystal structure of the isolated PhnH homodimer (Extended Data Fig. 5a)¹⁹. The CID is an insertion in PhnJ between $\beta 5$ and $\beta 6$ of the corresponding PhnH fold and consists of two α -helices and a short 3_{10} -helix (Fig. 2b). The CID is well conserved among PhnJ orthologues and contacts both of the central PhnI molecules (Fig. 3b). Finally, the CMD is located at the C terminus and consists of a small β -hairpin and a helix. It is stabilized by a zinc ion coordinated by four conserved cysteine residues: Cys241, Cys244, Cys266 and Cys272 (Fig. 2b).

The PhnI monomers bind each other via an extensive, conserved surface interaction area comprising $\sim 75\%$ of the total PhnGHIJ dimerization interface (Extended Data Fig. 6). Each molecule of PhnI interacts with both copies of PhnG (Fig. 3c), the smallest protein in the complex displaying an elongated $\alpha + \beta$ fold with a four-stranded, antiparallel β -sheet against a four-helix bundle (β -barrel domain, Fig. 2c). Despite very little sequence similarity, the closest known structural homologue of PhnG is PhnI, with which it shares both the long β -hairpin and the helical bundle (Fig. 2 and Extended Data Fig. 4b). The PhnG β -hairpin and C-terminal helix form a molecular clamp that connects to a groove in PhnI, forming an unusually long, combined β -barrel domain (80 Å, Fig. 3d).

The iron-sulfur binding site

PhnJ belongs to the anaerobic radical SAM enzyme superfamily in which three conserved cysteine residues coordinate a cubane-like Fe_4S_4 cluster²⁰ that promotes formation of a free electron radical required for catalysis by reductive cleavage of SAM to a 5'-deoxyadenosyl radical (Ado-CH_2^\bullet)

and L-methionine^{21–23}. PhnJ does not contain the canonical $\text{CX}_3\text{CX}_2\text{C}$ motif but rather a $\text{CX}_2\text{CX}_{21}\text{C}$ motif involving Cys241, Cys244 and Cys266, which are both necessary and sufficient for reconstitution of

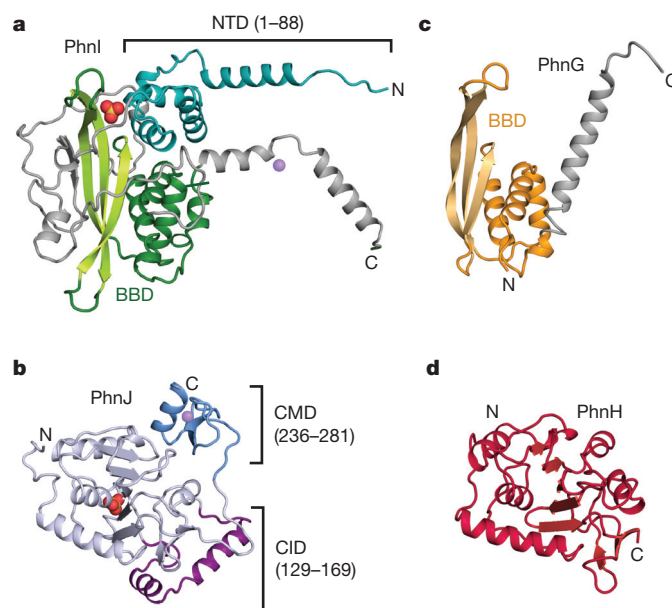


Figure 2 | Details of subunit structures. **a–d**, Details of the individual protein structures in the complex: PhnI (**a**), PhnJ (**b**), PhnG (**c**) and PhnH (**d**), aligned to show their structural homologies and with domain colours as in Fig. 1b. Ions are shown as spheres; sulfate, red and yellow; zinc, pink. BBD, β -barrel domain; CID, central insertion domain; CMD, C-terminal mini domain; NTD, N-terminal domain.

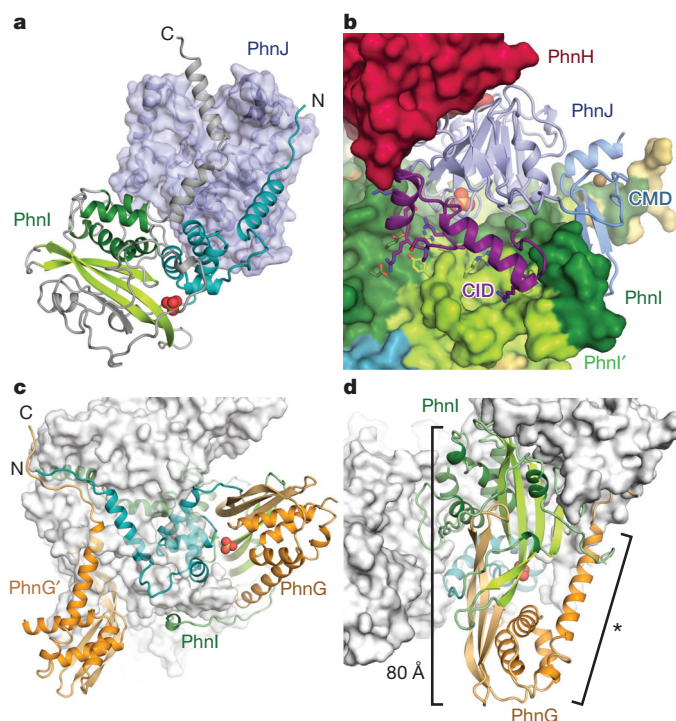


Figure 3 | Subunit interactions within the C-P lyase core complex. **a**, The termini of PhnI (grey, C and cyan, N) grasp PhnJ (semi-transparent blue surface). Domains in PhnI are coloured as in Fig. 1 and the sulfate ion is shown as spheres. **b**, The central insertion domain (CID, purple) contacts both copies of PhnI (green surfaces). Residues involved in the interaction are shown with sticks. **c**, PhnI (green/cyan) interacts with both copies of PhnG (orange). **d**, PhnG (orange) grabs both sides of PhnI (green) using the β -barrel domain and C-terminal helix (indicated with an asterisk). The combined β -barrel domain is 80 Å long.

the iron-sulfur cluster *in vitro*^{17,18}. In the structure, these cysteines are juxtaposed and coordinate a zinc ion (Fig. 4a) in an arrangement that closely matches that expected for an Fe_4S_4 cluster-containing protein (Fig. 4b). Furthermore, super-positioning an S-adenosyl methionine activase structure on this region reveals a small groove on the surface of the C-P lyase core complex next to the cluster site that might accommodate SAM²⁴ (Extended Data Fig. 5b).

According to the proposed reaction mechanism, the Ado-CH_2^\bullet radical is transferred to the universally conserved Gly32 of PhnJ, generating a stable glycyl radical enzyme that supports multiple turnovers without further SAM consumption¹⁷. According to this scheme, transfer of the radical from Gly32 to the fourth conserved cysteine residue (Cys272) generates a thiyl radical capable of homolytic C-P bond cleavage of 5-phosphoribosyl-1-phosphonate (Extended Data Fig. 1) through a thiophosphonate radical intermediate¹⁷. Cys272 is situated adjacent to the cluster site where it is the fourth ligand binding the zinc ion, while Gly32 is located more than 30 Å away, in the vicinity of PhnH (Fig. 4c). A direct involvement of Gly32 in the reaction²⁰ is therefore difficult to reconcile with the structure; however, it cannot be excluded that structural rearrangements could alter the position of the cluster relative to Gly32 to bring them into proximity. The CMD containing the cluster site has higher *B* factors, suggesting that it is relatively loosely attached to the PhnJ core and perhaps could detach during the reaction (Fig. 2b and Extended Data Fig. 5c).

A second potential active site

At the interface of PhnI and PhnJ, three universally conserved histidine residues come together to form a second metal-ion-binding site. Analysis of the anomalous difference density confirms that this His

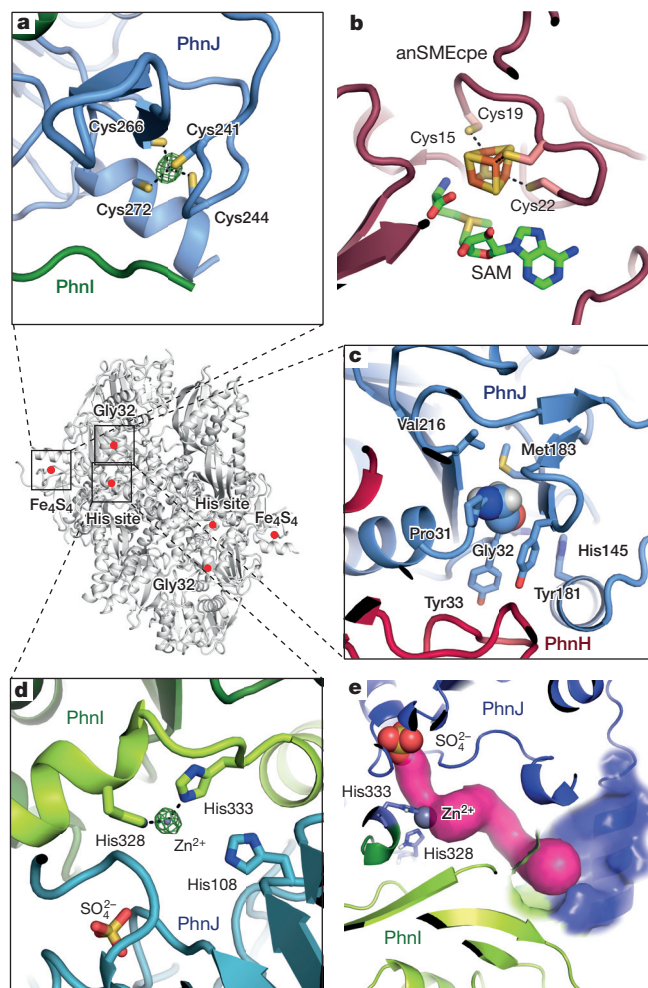


Figure 4 | Overview of the active sites of the C-P lyase core complex. **a**, Details of the iron-sulfur cluster site with a bound zinc ion (green, anomalous difference density at 6.0σ). **b**, The Fe_4S_4 cluster found in S-adenosyl methionine activase (anSMCpe, Protein Data Bank accession code: 4K37)²⁴ shown in the same orientation as in **a**. Sticks are coloured by standard atom colours (C, green/salmon; Fe, orange; S, yellow). **c**, The environment of PhnJ Gly32 (space fill). **d**, The zinc-ion-binding site at the interface between PhnI and PhnJ (His site) with the sulfate ion shown. **e**, A tunnel (magenta) leads from the sulfate (red and yellow) and zinc ions (grey) to the surface of the complex (blue area).

site also contains zinc (Fig. 4d). Two of the residues (PhnI His328 and His333) coordinate the zinc ion directly (2.4 Å), while the third (PhnJ His108) is further away (4.5 Å). The three histidines are located in a cavity between PhnI and PhnJ that connects to the surface of the complex via a solvent-accessible tunnel (Fig. 4e). The cavity also contains a sulfate ion located 9.5 Å from the zinc, which may mimic a substrate phosphate or phosphonate. Finally, access to the cavity is defined by the PhnJ CID domain, which forms a lid-like domain.

Studies of zinc-binding proteins show that structural zinc sites usually have four protein ligands, while active-site zinc ions have a more open coordination sphere with 2–3 ligands similar to that observed in this case²⁵. To assess the functional importance of the His site, we used genetic complementation to determine whether mutation of the histidine residues affects the ability of *E. coli* to utilize phosphonates. A plasmid-borne copy of the wild-type *phnGHJKLMN* allele was used to complement *E. coli* $\Delta\text{phn-HIJKLMN}$ under conditions where phosphonates were the sole phosphate source (Extended Data Fig. 7)²⁶. Unlike the wild type, none of the variants (PhnI(H333A), PhnI(H328A;H333A), PhnJ(H108A) and PhnJ(C272A)) could utilize phosphonates;

we therefore conclude that the His site is required for the activity of the C–P lyase core complex *in vivo*.

PhnK binds via the PhnJ CID

The C–P lyase core complex stably associates with a fifth component, PhnK (28 kDa)¹⁴. The function of PhnK is unclear but it contains the consensus elements of an ATP-binding cassette protein, suggesting that it might deliver nucleotides for the reaction (Extended Data Fig. 8)¹⁴. Despite its ability to stably co-purify, we were unable to obtain crystals of a complex including PhnK. We mapped the PhnK-binding site on the complex using negative-stain electron microscopy by generating a 3D reconstruction of purified PhnGHIJK (Fig. 5a and Extended Data Fig. 9). The crystal structure fits tightly within the resulting electron microscopy density map and reveals additional density in a groove close to the two-fold symmetry axis near two regions of highly conserved residues on PhnJ (Fig. 5a, b). The fold of PhnK can be roughly modelled using a homologous nucleotide-binding domain of an ABC transporter (Protein Data Bank accession code: 4FWI)²⁷. The electron microscopy map is consistent

with a single PhnK binding unilaterally to the complex, breaking the two-fold symmetry (Fig. 5c). Although the exact orientation of PhnK cannot be established at this resolution, we note that one side is highly conserved among orthologues, suggesting that it comprises the interaction surface (Fig. 5c).

ABC modules often dimerize in a head-to-tail fashion, binding ATP between the Walker A/B motifs of one subunit and the ABC motif of the other²⁸. PhnK contains a variant ABC motif (FSGGMQ versus LSGGQ), which could serve to bind the C–P lyase core complex (Extended Data Fig. 8). The conserved CID domain protrudes into the PhnK-binding region, so to probe its importance we constructed C–P lyase core complexes lacking residues 130–171 of PhnJ (Extended Data Fig. 10). Purification PhnGHIJ(Δ CID)K demonstrated that upon deletion of the CID, the C–P lyase core complex remains intact but PhnK is missing, thus indicating that the CID region of PhnJ is required for tethering PhnK to the core complex.

Discussion

In this paper, we delineate the organization and detailed molecular structure of a core complex involved in phosphonate catabolism in bacteria. We show that four of the proteins required for phosphonate breakdown assemble into a large, hetero-octameric core complex with two-fold symmetry and that the symmetry is broken by binding of a fifth, ATP-binding subunit, PhnK. The structure is not immediately compatible with the direct involvement of Gly32 (PhnJ) in catalysis, but structural rearrangements may affect the location of this residue during the reaction. Many glycyl radical enzymes require separate activation enzymes that dissociate upon radical formation, a task that could also be maintained by a flexible internal domain²⁰.

The structure indicates the existence of a second active site at the interface of PhnI and PhnJ. Analysis of difference electron density maps from several independent data sets revealed a consistent density next to the bound zinc ion, but we have been unable to identify the bound molecule. We also carried out co-crystallization using a range of compounds including nucleotides and phosphonates, but no further substrate binding was observed. This suggests that the complex needs an Fe₄S₄ cluster, or is not in the correct conformational state to bind a substrate. We speculate that the His site is required for coupling a phosphonate to ATP, which is known to depend on PhnI¹⁸.

Using electron microscopy we locate the binding region for PhnK on the C–P lyase core complex. While this does not reveal the role of PhnK in the reaction, we note that the region is close to the His site and it is therefore possible that structural changes in PhnK occurring upon ATP hydrolysis may affect substrate access. With the detailed architecture of the C–P lyase core complex thus delineated, future work will focus on understanding the requirements of the two reactions catalysed by the complex and definitively locating the binding sites of substrates and reaction intermediates.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 7 January; accepted 22 June 2015.

Published online 17 August 2015.

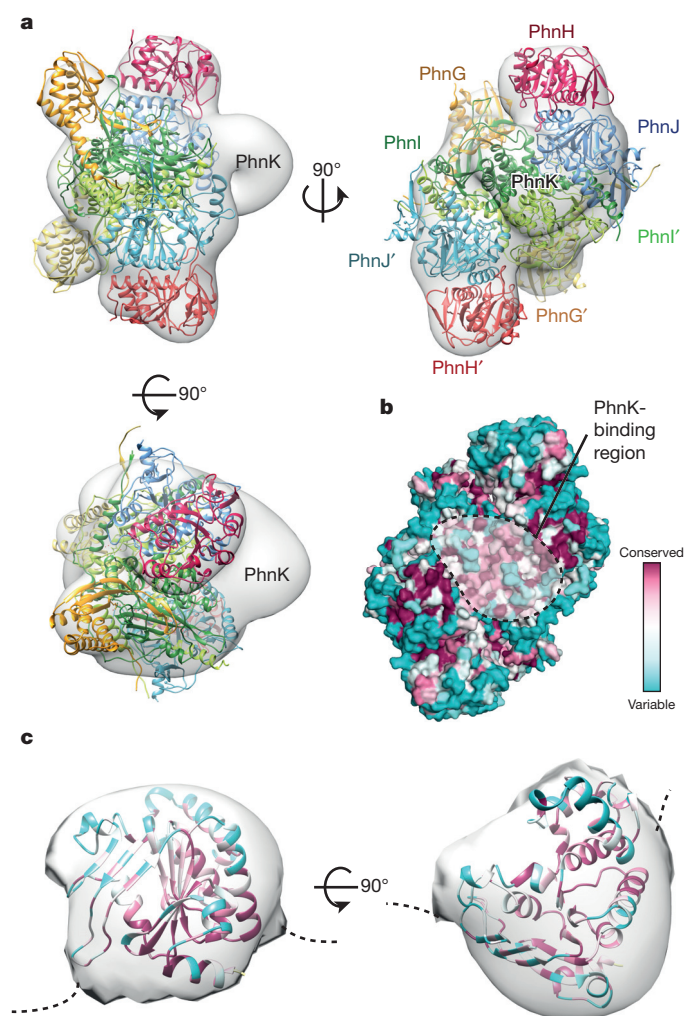


Figure 5 | Mapping of PhnK using electron microscopy. **a**, Orthogonal views of the negative stain electron microscopy 3D reconstruction with the C–P lyase core complex crystal structure docked and the position of PhnK indicated. Colours are as in Fig. 1c. **b**, Surface conservation of the C–P lyase core complex from variable (teal) to conserved (burgundy) with the PhnK-binding region (dashed area) indicated. **c**, A single PhnK molecule (based on Protein Data Bank entry 4FWI)²⁷ fitted into the electron microscopy density masked to remove the C–P lyase core complex. The model is coloured by conservation using ConSurf²⁹ at the same scale as in **b**.

- Karl, D. M. Aquatic ecology: Phosphorus, the staff of life. *Nature* **406**, 31–33 (2000).
- Hsieh, Y. J. & Wanner, B. L. Global regulation by the seven-component Pi signaling system. *Curr. Opin. Microbiol.* **13**, 198–203 (2010).
- Hove-Jensen, B., Zechel, D. L. & Jochimsen, B. Utilization of glyphosate as phosphorus source: biochemistry and genetics of bacterial carbon-phosphorus lyase. *Microbiol. Mol. Biol. Rev.* **78**, 176–197 (2014).
- Imazu, K. *et al.* Enhanced utilization of phosphonate and phosphite by *Klebsiella aerogenes*. *Appl. Environ. Microbiol.* **64**, 3754–3758 (1998).
- Quinn, J. P. Carbon phosphorus lyase activity — a novel mechanism of bacterial resistance to the phosphonic acid antibiotics. *Lett. Appl. Microbiol.* **8**, 113–116 (1989).
- Schowaneck, D. & Verstraete, W. Phosphonate utilization by bacterial cultures and enrichments from environmental samples. *Appl. Environ. Microbiol.* **56**, 895–903 (1990).

7. Wackett, L. P., Shames, S. L., Venditti, C. P. & Walsh, C. T. Bacterial carbon-phosphorus lyase: products, rates, and regulation of phosphonic and phosphinic acid metabolism. *J. Bacteriol.* **169**, 710–717 (1987).
8. White, A. K. & Metcalf, W. W. Microbial metabolism of reduced phosphorus compounds. *Annu. Rev. Microbiol.* **61**, 379–400 (2007).
9. Chen, C. M., Ye, Q. Z., Zhu, Z. M., Wanner, B. L. & Walsh, C. T. Molecular biology of carbon-phosphorus bond cleavage. Cloning and sequencing of the *phn* (*psiD*) genes involved in alkylphosphonate uptake and C-P lyase activity in *Escherichia coli* B. *J. Biol. Chem.* **265**, 4461–4471 (1990).
10. Hove-Jensen, B., McSorley, F. R. & Zechel, D. L. Physiological role of *phnP*-specified phosphoribosyl cyclic phosphodiesterase in catabolism of organophosphonic acids by the carbon-phosphorus lyase pathway. *J. Am. Chem. Soc.* **133**, 3617–3624 (2011).
11. Hove-Jensen, B., Rosenkrantz, T. J., Haldimann, A. & Wanner, B. L. *Escherichia coli phnN*, encoding ribose 1,5-bisphosphokinase activity (phosphoribosyl diphosphate forming): dual role in phosphonate degradation and NAD biosynthesis pathways. *J. Bacteriol.* **185**, 2793–2801 (2003).
12. Metcalf, W. W. & Wanner, B. L. Evidence for a fourteen-gene, *phnC* to *phnP* locus for phosphonate metabolism in *Escherichia coli*. *Gene* **129**, 27–32 (1993).
13. Rizk, S. S., Cuneo, M. J. & Hellinga, H. W. Identification of cognate ligands for the *Escherichia coli phnD* protein product and engineering of a reagentless fluorescent biosensor for phosphonates. *Protein Sci.* **15**, 1745–1751 (2006).
14. Jochimsen, B. *et al.* Five phosphonate operon gene products as components of a multi-subunit complex of the carbon-phosphorus lyase pathway. *Proc. Natl Acad. Sci. USA* **108**, 11393–11398 (2011).
15. Ren, Z. *et al.* Subunit interactions within the carbon-phosphorus lyase complex from *Escherichia coli*. *Biochemistry* **54**, 3400–3411 (2015).
16. Frost, J. W., Loo, S., Cordeiro, M. L. & Li, D. Radical-based dephosphorylation and organophosphonate biodegradation. *J. Am. Chem. Soc.* **109**, 2166–2171 (1987).
17. Kamat, S. S., Williams, H. J., Dangott, L. J., Chakrabarti, M. & Raushel, F. M. The catalytic mechanism for aerobic formation of methane by bacteria. *Nature* **497**, 132–136 (2013).
18. Kamat, S. S., Williams, H. J. & Raushel, F. M. Intermediates in the transformation of phosphonates to phosphate by bacteria. *Nature* **480**, 570–573 (2011).
19. Adams, M. A. *et al.* Crystal structure of PhnH: an essential component of carbon-phosphorus lyase in *Escherichia coli*. *J. Bacteriol.* **190**, 1072–1083 (2008).
20. Shisler, K. A. & Broderick, J. B. Glycyl radical activating enzymes: structure, mechanism, and substrate interactions. *Arch. Biochem. Biophys.* **546**, 64–71 (2014).
21. Booker, S. J. & Grove, T. L. Mechanistic and functional versatility of radical SAM enzymes. *F1000 Biol. Rep.* **2**, 52 (2010).
22. Frey, P. A., Hegeman, A. D. & Ruzicka, F. J. The radical SAM superfamily. *Crit. Rev. Biochem. Mol. Biol.* **43**, 63–88 (2008).
23. Sofia, H. J., Chen, G., Hetzler, B. G., Reyes-Spindola, J. F. & Miller, N. E. Radical SAM, a novel protein superfamily linking unresolved steps in familiar biosynthetic pathways with radical mechanisms: functional characterization using new analysis and information visualization methods. *Nucleic Acids Res.* **29**, 1097–1106 (2001).
24. Goldman, P. J. *et al.* X-ray structure of an AdoMet radical activase reveals an anaerobic solution for formylglycine posttranslational modification. *Proc. Natl Acad. Sci. USA* **110**, 8519–8524 (2013).
25. McCall, K. A., Huang, C. & Fierke, C. A. Function and mechanism of zinc metalloenzymes. *J. Nutr.* **130**, 1437S–1446S (2000).
26. Yakovleva, G. M., Kim, S. K. & Wanner, B. L. Phosphate-independent expression of the carbon-phosphorus lyase activity of *Escherichia coli*. *Appl. Microbiol. Biotechnol.* **49**, 573–578 (1998).
27. Li, X. *et al.* Structure of the nucleotide-binding domain of a dipeptide ABC transporter reveals a novel iron-sulfur cluster-binding domain. *Acta Crystallogr. D* **69**, 256–265 (2013).
28. Hollenstein, K., Dawson, R. J. & Locher, K. P. Structure and mechanism of ABC transporter proteins. *Curr. Opin. Struct. Biol.* **17**, 412–418 (2007).
29. Landau, M. *et al.* ConSurf 2005: the projection of evolutionary conservation scores of residues on protein structures. *Nucleic Acids Res.* **33**, W299–W302 (2005).

Acknowledgements We are thankful to T. L. Sørensen at Diamond, T. Weinert at SLS as well as beamline staff at ESRF and MAX-Lab for help during X-ray data acquisition and S. Chen, C. G. Savva, J. Grimmert and T. Darling at the MRC-LMB for technical assistance with electron microscopy. This work was supported by the European Research Council grant no. 261151 (L.A.P.), MRC grant MC_U105192715 (L.A.P.), and the Danish National Research Foundation 'Centre for mRNP biogenesis and metabolism' (D.E.B.).

Author Contributions P.S., L.A.P., B.H.J., B.J. and D.E.B. designed and P.S., L.B.V., C.J.R. and B.J. carried out the experiments. P.S., M.K. and D.E.B. determined the crystal and EM structures while C.J.R. and L.A.P. carried out final refinement of the EM structure as well as EM structure validation. P.S., M.K., C.J.R., L.A.P., B.H.J., B.J. and D.E.B. wrote the manuscript.

Author Information Atomic coordinates and structure factors have been deposited in the Protein Data Bank (PDB) with accession code 4XB6. The EM density map has been deposited in the Electron Microscopy Data Bank (EMDB) with accession code EMD-3033. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.E.B. (deb@mbg.au.dk).

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Protein expression and protein purification. The construction of pHO572 (expressing *phnGHII*) and pHO575 (expressing *phnGHIIK*) as well as gene expression in *E. coli* strain HO2735 ($\Delta(lac)X74 \Delta(phnCDEFGHIJLMNOP)$ 33–30/F *lacI^q* *zzf::Tn10*) were described previously¹⁴. pHO575 encodes a C-terminally six-histidine tagged version of PhnK while PhnGHII has no tag but still binds to Ni²⁺ NTA agarose beads. The PhnGHII(ACID) and PhnGHII(ACID)K constructs were created by site-directed mutagenesis using primers 5'-GTGCCAATCCCCGAGGGCGGCTATCCGGTGAAGGTA-3' (ACID forward), and 5'-TACCTTACCCGATAGCCGCCCTCGGGGATTGGCAC-3' (ACID reverse) which result in the replacement of residues 130–172 of PhnJ by two glycine residues (underlined in the primers). Cells were in all cases grown at 37 °C in Luria Broth (LB) medium and gene expression achieved overnight at 18 °C by induction using 0.5 mM isopropyl- β -D-thiogalactoside (IPTG). Cells were pelleted by centrifugation at 16,000 r.p.m. for 45 min at 4 °C and resuspended in lysis buffer (50 mM HEPES, pH 7.5, 500 mM NaCl, 5 mM MgCl₂, 20% (v/v) glycerol, and 3 mM 2-mercaptoethanol) supplemented with Complete Protease Inhibitor Cocktail tablets (Sigma) and lysed by high-pressure homogenization (EmulsiFlex-C5, Evesin) at 15,000 p.s.i. The lysed cells were centrifuged at 16,000 r.p.m. for 45 min to remove cell debris and bound to Ni²⁺ NTA agarose beads on a 5-ml pre-packed HisTrap HP column (GE Healthcare), pre-equilibrated with lysis buffer (PhnGHII) or lysis buffer plus 20 mM imidazole (PhnGHIIK). In all cases, the complexes were eluted by increasing the imidazole concentration to 250 mM. Following overnight dialysis at 4 °C against buffer LS1 (50 mM HEPES, pH 7.5, 100 mM NaCl, 5 mM MgCl₂, and 5 mM 2-mercaptoethanol), the samples were applied to a 1 ml Source 15Q column (GE Healthcare), pre-equilibrated with buffer LS1 and eluted using a linear gradient from 100–600 mM NaCl. The samples were then diluted to reach 250 mM NaCl and passed over a 1 ml Mono Q column (GE Healthcare) pre-equilibrated with buffer LS2 (50 mM HEPES, pH 7.5, 250 mM NaCl, 5 mM MgCl₂, and 5 mM 2-mercaptoethanol), washed, and eluted using a 250–400 mM NaCl gradient. Finally, the complexes were purified on a Superdex 200 10/300 GL size-exclusion column (GE Healthcare) equilibrated with buffer GF (50 mM HEPES, pH 7.5, 300 mM NaCl, and 5 mM 2-mercaptoethanol). A 2 l culture typically yielded 4–8 mg of a purified protein complex. Purification from the PhnGHII(ACID) and PhnGHII(ACID)K constructs was stopped after the Source Q column.

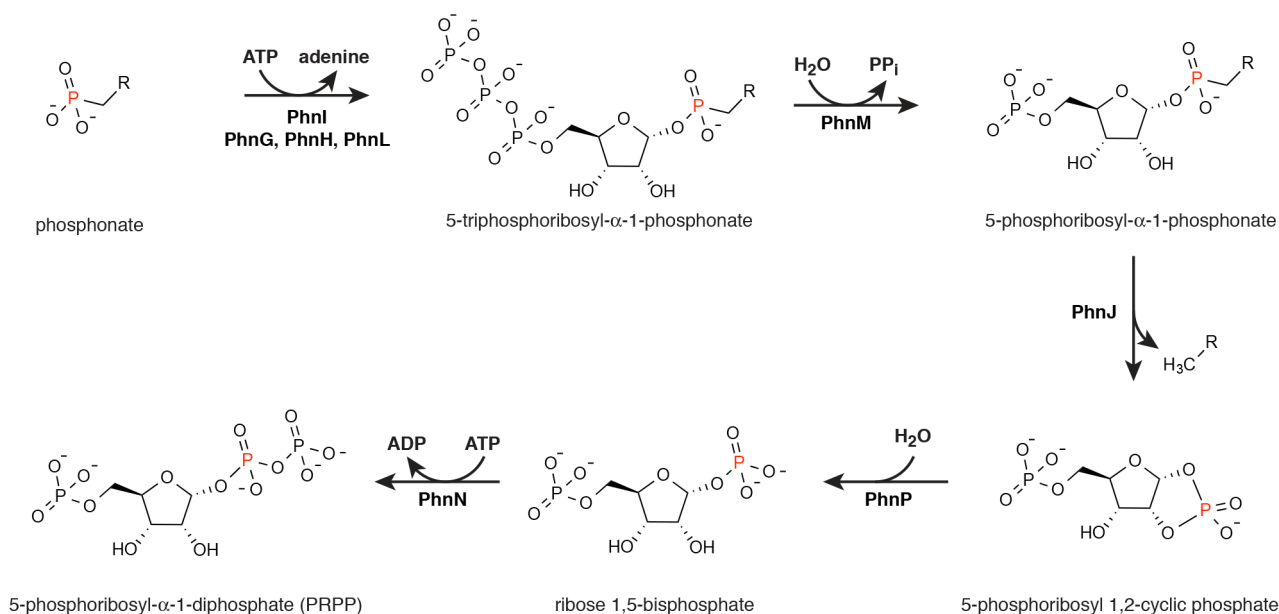
Crystallization and structure determination of PhnGHII. Crystals of the PhnGHII complex were obtained using batch crystallization at 4 °C with a reservoir solution containing 20% (w/v) PEG 10,000, 0.1 M HEPES, pH 7.5, 1 M trisodium citrate dihydrate, 2.3% (w/v) 1,8-diaminooctane, and 5 mM 2-mercaptoethanol. Crystallization drops contained 1 μ l protein sample mixed with 0.7 μ l reservoir solution and 0.3 μ l of a micro-seed stock³⁰ obtained from early stage hits. Crystals suitable for data collection appeared within 2–3 days reaching maximum dimensions of 1.0 \times 0.2 \times 0.2 mm. Crystals were collected, cryo-protected by gradual addition of glycerol to a final concentration of 25% (v/v), and flash-frozen in liquid nitrogen. For structure solution by molecular replacement with single-wavelength anomalous dispersion (MR-SAD), crystals were derivatized for 24 h with Ta₆Br₁₂ (Proteros Biostructures)³¹, which was added directly to the crystallization drop as powder. Diffraction data were collected at 100 K at the X06DA beamline at SLS, Villigen, Switzerland on a PILATUS 2M detector (native crystals) or a PILATUS 6M detector at the X06SA beamline (Ta₆Br₁₂ data). Data sets were processed with xia2 (ref. 32) and the structure was solved by MR-SAD method using Phaser and AutoSol via PHENIX³³. The PhnH structure¹⁹ (PDB ID: 2FSU) was used as a molecular replacement search model in Phaser to locate the Ta₆Br₁₂ sites by MR-SAD, whereby the partial molecular replacement solution allowed identification of the Ta₆Br₁₂ cluster sites and served as a source of phase information in phenix.autosol. Initial phases were obtained after density modification using RESOLVE in the phenix.autosol pipeline and the resulting maps used to auto-build secondary structure elements using ARP/wARP from the CCP4 package³⁴. The resulting partial model was then used as a search model for molecular replacement against the native data using Phaser and run through 25 cycles of backbone auto-tracing using the native data set in SHELXE³⁵. Finally, a near-complete model of PhnGHII could be auto-built using phenix.autobuild, RESOLVE, and Buccaneer³⁶. Missing parts of the model were completed manually in Coot³⁷. The model was iteratively improved and refined using phenix.refine³³, and validated using MolProbity³⁸. The final model contains 1,958 amino acid residues, four bound sulfate ions, four zinc ions, and 1,792 molecules of ordered solvent. The model is complete except for 18 residues exclusively located at the termini of the subunits. Structure figures were made using PyMOL³⁹, Chimera⁴⁰, and ConSurf²⁹.

Electron microscopy. Purified samples of PhnGHIIK samples were applied to Quantifoil R2/2 holey carbon on copper grids (Quantifoil, Jena)⁴¹, covered with an additional thin film of amorphous carbon, and rendered more hydrophilic with a 9:1 argon–oxygen plasma (Fischione Model 1070). The specimens were stained with 3% ammonium molybdate at pH 8 followed by 2% uranyl acetate. Micrographs were recorded at 44,000 \times magnification on a Tecnai T12 microscope equipped with a US4000 4K \times 4K pixel CCD detector (Gatan) at 120 keV with defoci in the 0.8–2 μ m range and using an electron dose of 20 electrons/ \AA^2 . 10,137 single particles were manually picked from 105 micrographs using e2boxer from the EMAN2 software package⁴² and contrast transfer function parameters were determined using CTFFIND3 (ref. 43). Three iterations of 2D classification were performed in RELION⁴⁴ using 300 class averages to determine particles that did not align well with each other. These particles were removed from subsequent analysis. After 2D classification, the final set of 10,033 particles was used to calculate a 3D reconstruction in RELION without symmetry imposed. The initial model for the reconstruction was prepared by low-pass filtering a density map generated from the C–P lyase core complex crystal structure to 40 \AA . The final model has a resolution of 16 \AA by the 0.143 'gold standard' Fourier-shell correlation^{45,46} and a resolution of 28 \AA versus the crystal structure at FSC = 0.5. The latter is probably closer to the true resolution of the map as the granular nature of negative stain can introduce correlations in the half-maps that are not related to the protein structure. The map was validated using 419 tilt-pairs recorded using angles 0° and 30° ($P \leq 0.01$, $\kappa = 2.7$)⁴⁷. The FSC versus the crystal structure shows correlation between the crystal structure and the electron microscopy density at low resolution, after which deviations due to structural differences between the C–P lyase core complex and the PhnGHIIK complex become apparent.

In vivo complementation. For the *in vivo* complementation studies, *E. coli* strain BW16711 ($\Delta phnHIIJLMNOP$) was transformed by the plasmid pGY1²⁶, conferring ampicillin resistance and encoding *phnGHIIJLMNOP*, and analysed for its ability to grow on MOPS minimal plates⁴⁸ supplemented with 0.2% glucose, 100 μ g ml⁻¹ ampicillin, 0.1 mM IPTG and 0.2 mM of either methyl phosphonate, 2-aminoethyl phosphonate, or phosphate ion as a positive control. The PhnI(H328A) variant, the PhnI(H328A;H333A) double variant, the PhnJ(C272A) variant as well as the PhnJ(H108A) variant, were introduced into pGY1 by site-directed mutagenesis by PCR using the following primers, 5'-GCAGGCTTTGTCTCGGCCCTCAAACCTC CCCC-3' (H328A forward), 5'-TGGGGGAGTTTGAAGGCCGAGACAAAG CCTGC-3' (H328A reverse), 5'-GCAGGCTTTGTCTCGGCCCTCAAACCTCC CGCCTACGTCGATTTC-3' (H328A;H333A forward), 5'-TGGAAATCGAC GTAGGCGGGGAGTTTGAAGGCCGAGACAAAGCCTGC-3' (H328A;H333A reverse), 5'-TCCGATACCGATTATGCCGCCAACAGAGCGA-3' (C272A forward), 5'-TCGCTCTGTTGGCGGCGCATATCGGTATCGGA-3' (C272A reverse), 5'-CTTATCCAGACGCGTGGCCGCATCCCCGAAAC-3' (H108A forward), and 5'-GTTTCGGGGATGCGGGCAGCGTCTGGATAAG-3' (H108A reverse), where changes relative to the wild-type sequence are underlined. Template DNA was digested by the methylation-dependent endonuclease, DpnI, before transformation of non-ligated DNA into NovaBlue Singles (Novagen) electrocompetent *E. coli* cells and selection on ampicillin plates. All mutations were confirmed by sequencing of the entire *phnGHIIJLMNOP* region of the pGY1 vector to ensure that no other spontaneous mutations had been introduced that could prevent rescue of the BW16711 strain.

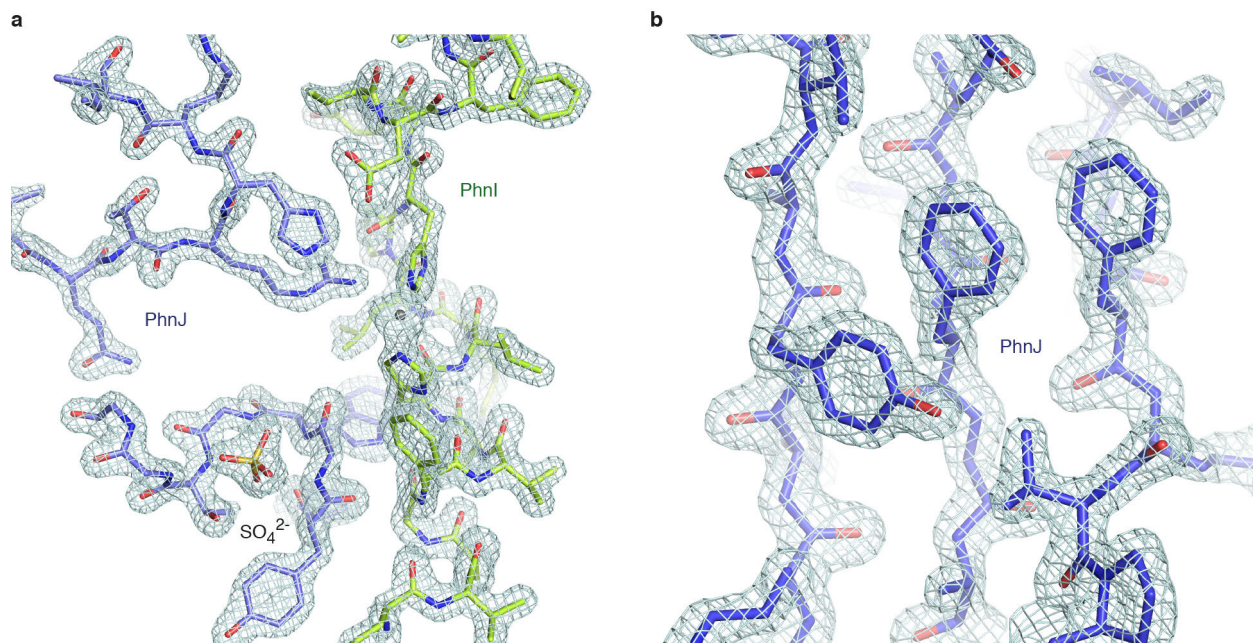
- D'Arcy, A., Villard, F. & Marsh, M. An automated microseed matrix-screening method for protein crystallization. *Acta Crystallogr. D* **63**, 550–554 (2007).
- Knäblein, J. et al. Ta₆Br₁₂²⁺, a tool for phase determination of large biological assemblies by X-ray crystallography. *J. Mol. Biol.* **270**, 1–7 (1997).
- Winter, G. xia2: an expert system for macromolecular crystallography data reduction. *J. Appl. Cryst.* **43**, 186–190 (2010).
- Adams, P. D. et al. PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. D* **66**, 213–221 (2010).
- Cohen, S. X. et al. Towards complete validated models in the next generation of ARP/wARP. *Acta Crystallogr. D* **60**, 2222–2229 (2004).
- Thorn, A. & Sheldrick, G. M. Extending molecular-replacement solutions with SHELXE. *Acta Crystallogr. D* **69**, 2251–2256 (2013).
- Cowan, K. The Buccaneer software for automated model building. 1. Tracing protein chains. *Acta Crystallogr. D* **62**, 1002–1011 (2006).
- Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D* **66**, 486–501 (2010).
- Chen, V. B. et al. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66**, 12–21 (2010).
- Schrödinger, LLC. The PyMOL molecular graphics system, version 1.3r1. (2010).
- Pettersen, E. F. et al. UCSF Chimera—a visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612 (2004).
- Ermantraut, E., Wohlfart, K. & Tichelaar, W. Perforated support foils with pre-defined hole size, shape and arrangement. *Ultramicroscopy* **74**, 75–81 (1998).
- Tang, G. et al. EMAN2: an extensible image processing suite for electron microscopy. *J. Struct. Biol.* **157**, 38–46 (2007).

43. Mindell, J. A. & Grigorieff, N. Accurate determination of local defocus and specimen tilt in electron microscopy. *J. Struct. Biol.* **142**, 334–347 (2003).
44. Scheres, S. H. RELION: implementation of a Bayesian approach to cryo-EM structure determination. *J. Struct. Biol.* **180**, 519–530 (2012).
45. Henderson, R. *et al.* Outcome of the first electron microscopy validation task force meeting. *Structure* **20**, 205–214 (2012).
46. Scheres, S. H. & Chen, S. Prevention of overfitting in cryo-EM structure determination. *Nature Methods* **9**, 853–854 (2012).
47. Russo, C. J. & Passmore, L. A. Robust evaluation of 3D electron cryomicroscopy data using tilt-pairs. *J. Struct. Biol.* **187**, 112–118 (2014).
48. Neidhardt, F. C., Bloch, P. L. & Smith, D. F. Culture medium for enterobacteria. *J. Bacteriol.* **119**, 736–747 (1974).



Extended Data Figure 1 | The conversion of a phosphonate to 5-phosphoribosyl-α-1-diphosphate (PRPP). PhnI supported by PhnG, PhnH and PhnL catalyses the transfer of the phosphonate moiety to the 1' position of the ribose of ATP through displacement of adenine, generating a 5-triphosphoribosyl-α-1-phosphonate. Subsequent to the removal of pyrophosphate by PhnM yielding a 5-phosphoribosyl-α-1-phosphonate, PhnJ

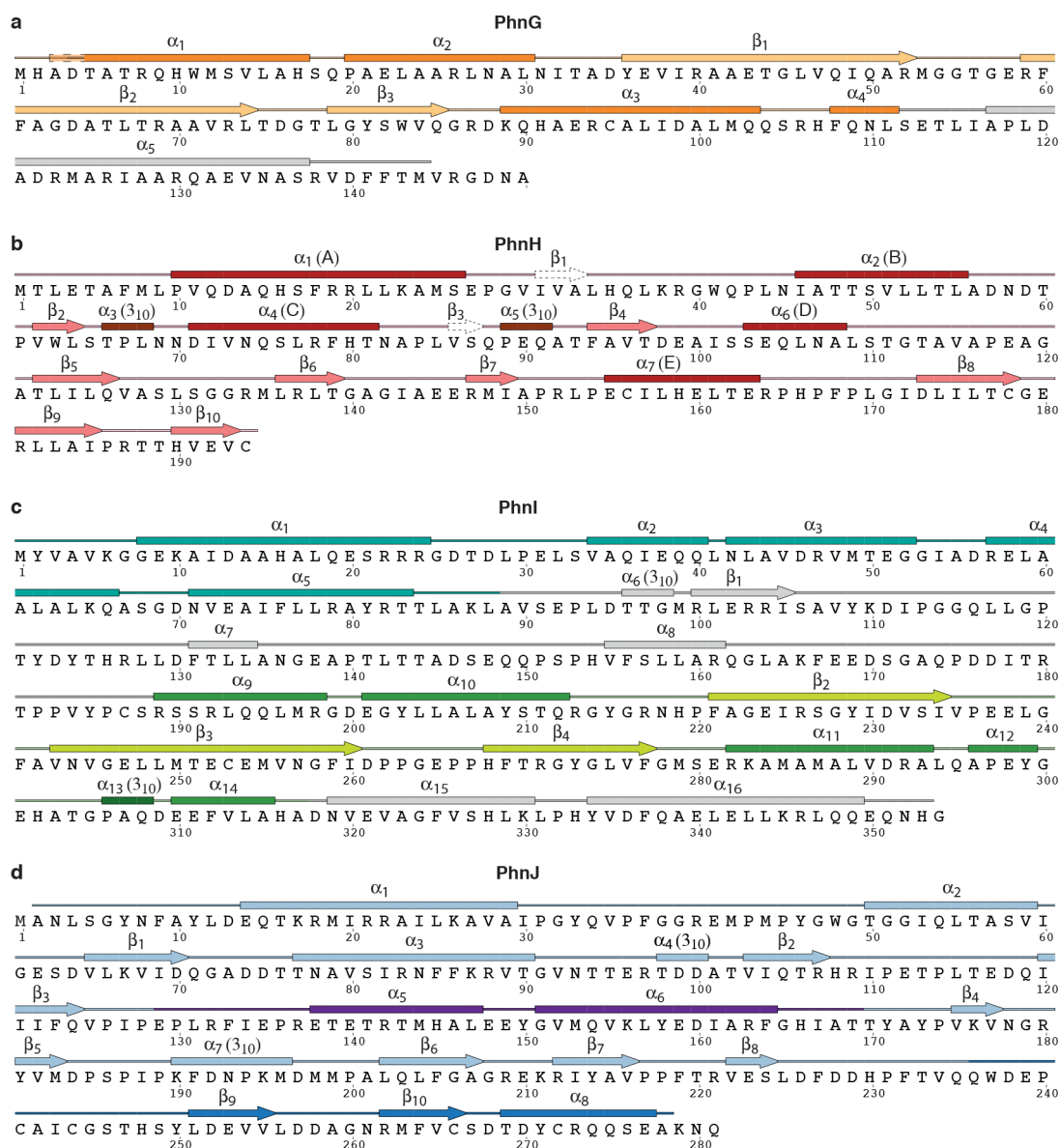
breaks the C-P bond of the ribose-coupled phosphonate liberating the alkyl moiety and generating 5-phosphoribosyl 1,2-cyclic phosphate. Finally, the combined activities of PhnP (a phosphoribosyl cyclic phosphodiesterase) and PhnN (a ribosyl bisphosphate phosphokinase) result in the formation of PRPP via ribose 1,5-bisphosphate.



Extended Data Figure 2 | Representative examples of electron density.

a, The interface between PhnJ (blue, residues 45–52 and 104–110 including a bound sulfate ion) and PhnI (green, residues 321–341) showing $2F_o - F_c$

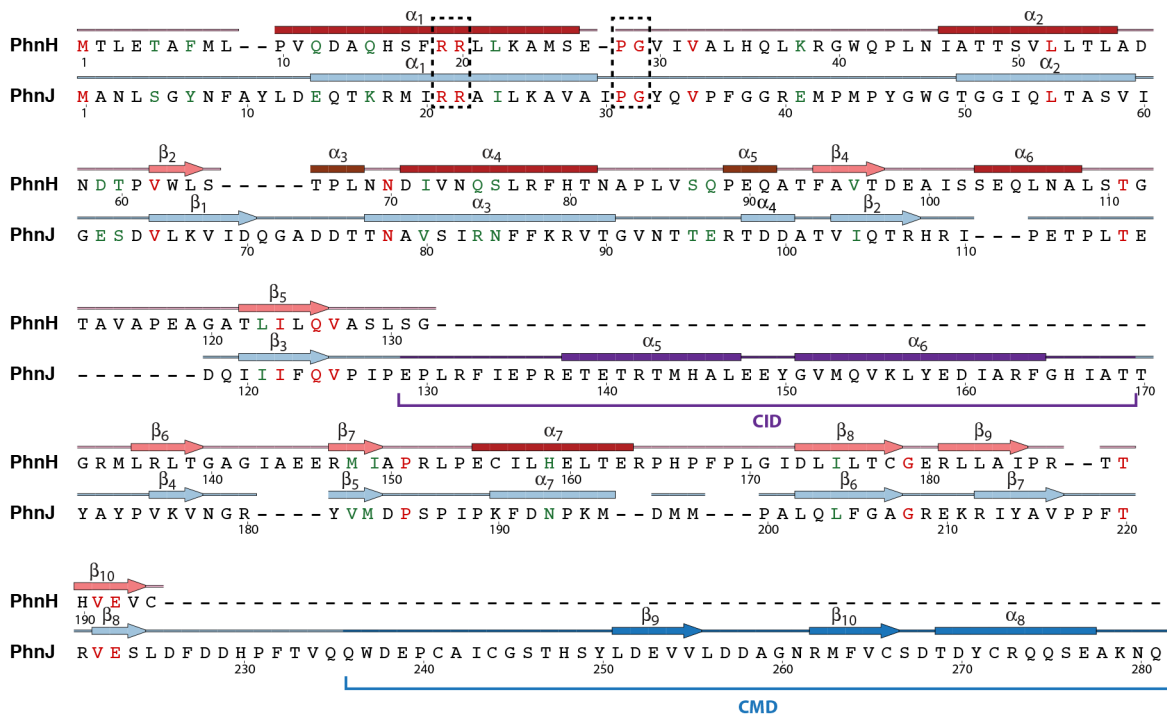
electron density contoured at 2.0 r.m.s.d. **b**, Close-up of the aromatic side chains in the central β -sheet of PhnJ (residues 118–126, 203–207 and 211–217), with the same contouring as **a**.



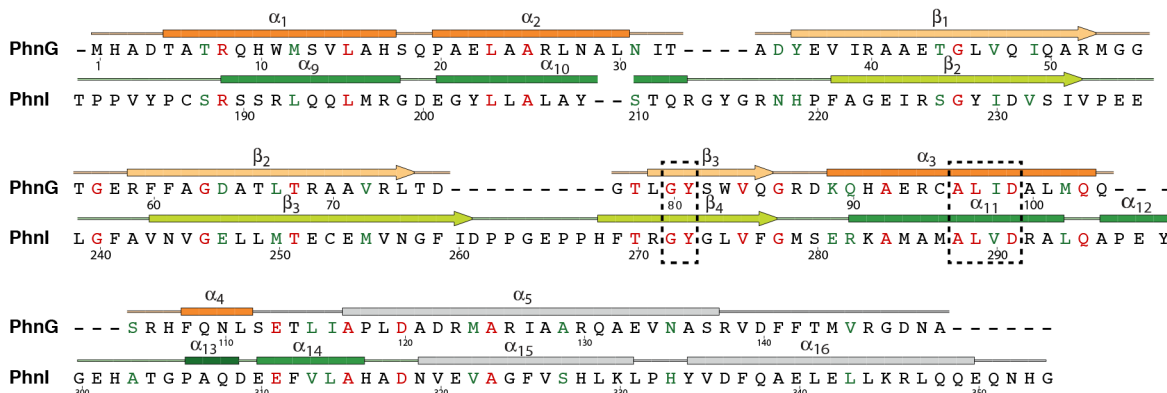
Extended Data Figure 3 | Sequences of the proteins of the C-P lyase core complex with secondary structure. Protein sequences are shown along with secondary structure assignment based on the crystal structure and colours as in Fig. 1. **a**, PhnG. The first α-helix is two residues longer in one of the two copies in the complex and indicated with a dashed box. The β-barrel domain is shown in yellow and orange colours. **b**, PhnH. The numbering of β-strands follows the convention from ref. 19. Helix names (A–E) used in that paper are shown in

parentheses. β₁ and β₃ are not included in the figures in this paper as they only have two hydrogen bonds each. **c**, PhnI. The N-terminal domain is shown with sea green and the β-barrel domain with green and light-green colours. **d**, PhnJ. The central insertion domain is shown in purple and the C-terminal mini domain in a darker blue colour. Figure produced using SecSeq (D. E. Brodersen, unpublished, <http://www.bioxray.au.dk/~deb/secseq>).

a

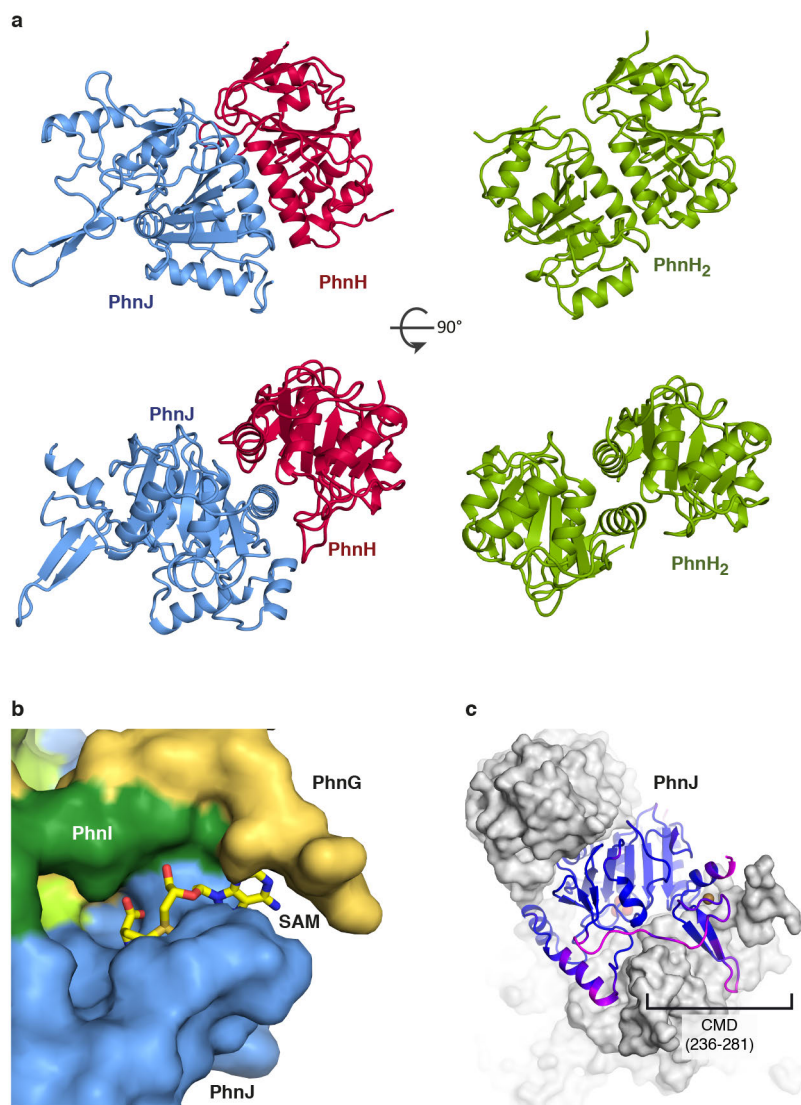


b



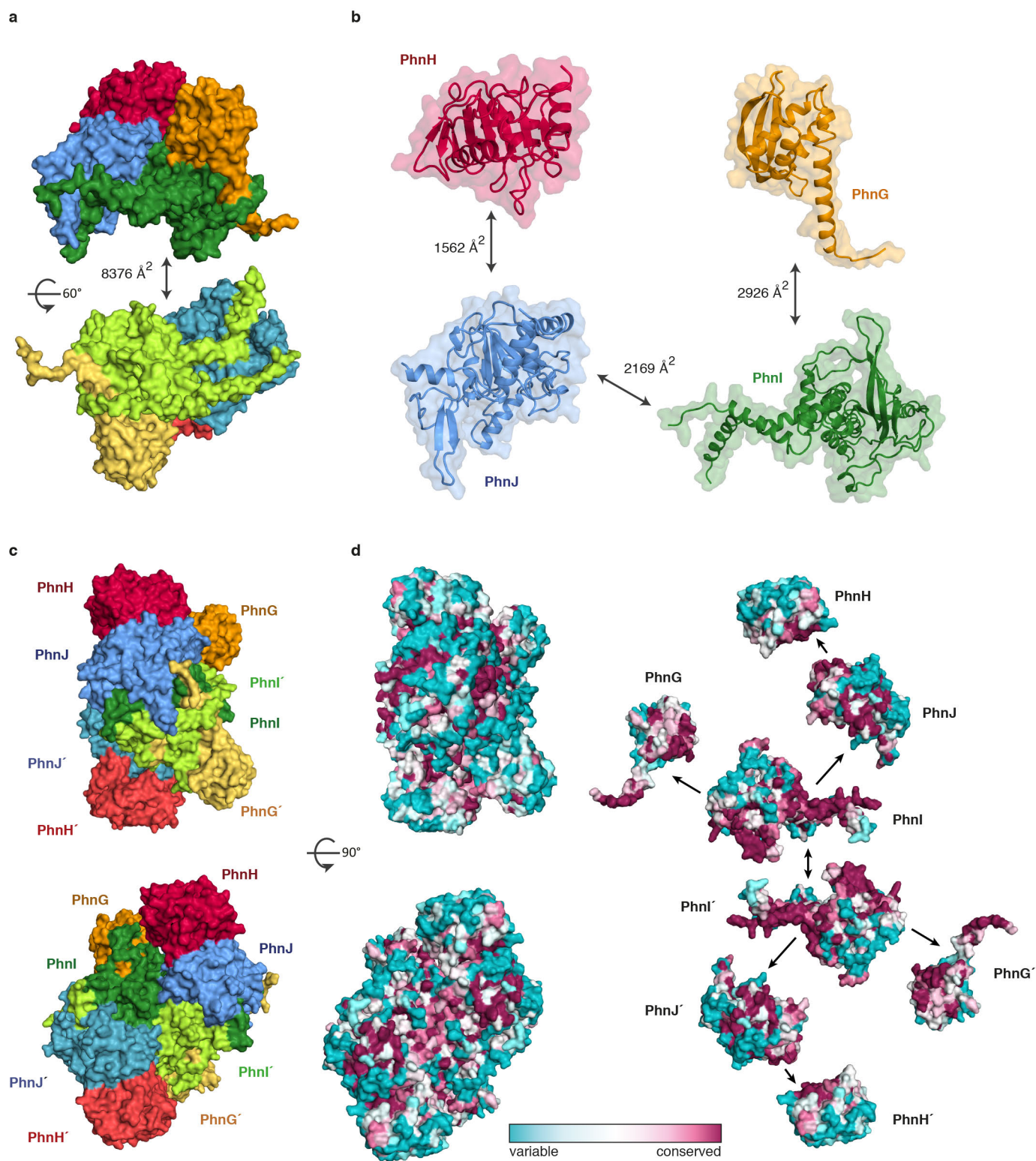
Extended Data Figure 4 | Cross alignments. **a**, Alignment of the amino acid sequences of *E. coli* PhnH and PhnJ. Identical residues are shown in red and conserved functionality in green. Secondary structure colours correspond to Fig. 1 and conserved regions are shown in dashed boxes. For PhnJ, the positions

of the CID and CMD are indicated with brackets as well as with colours. **b**, Alignment of PhnG and PhnI (only partial sequence). Conserved regions are shown in dashed boxes. Figure produced using SecSeq (D. E. Brodersen, unpublished, <http://www.bioxray.au.dk/~deb/secseq>).



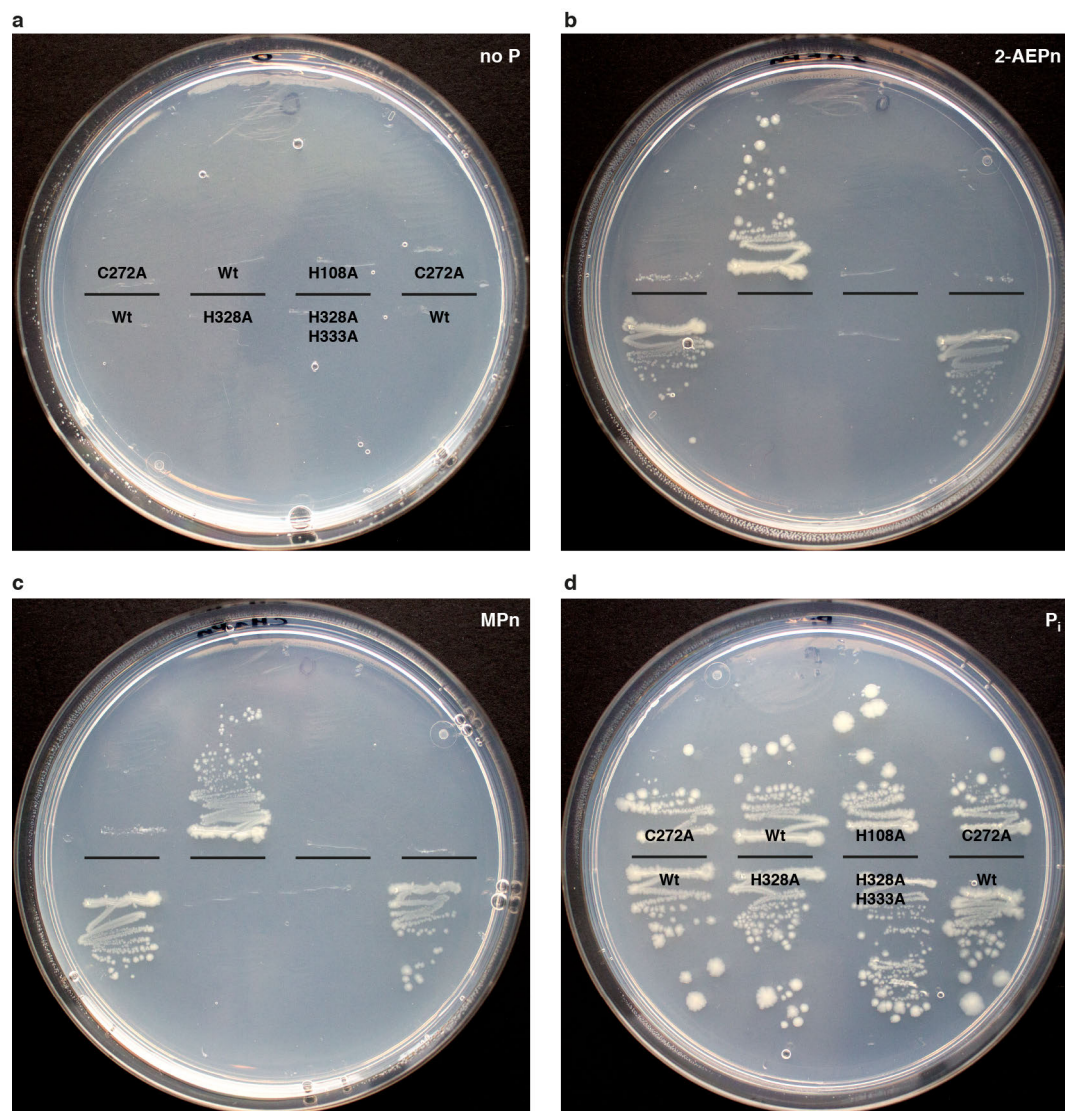
Extended Data Figure 5 | The structure and function of PhnJ. **a**, Two perpendicular views of the PhnH–PhnJ heterodimer as observed within the C–P lyase core complex (blue and red, left) and aligned views of the PhnH₂ homodimer from the isolated crystal structure (PDB ID: 2FSU; green, right)¹⁹. **b**, Surface view of the C–P lyase core complex with PhnG shown in yellow, PhnI in green, and PhnJ in blue. The position of SAM as modelled from an

S-adenosyl methionine activase enzyme (PDB ID: 4K37)²⁴ has been overlaid to visualize its putative placement in the pocket between PhnG and PhnJ. **c**, The C–P lyase core complex shown in surface representation with PhnJ and the CMD in cartoon, coloured by *B* factor to show flexibility ($B = 25 \text{ \AA}^2$, blue, $B = 45 \text{ \AA}^2$, magenta). The zinc and sulfate ions are shown with spheres.



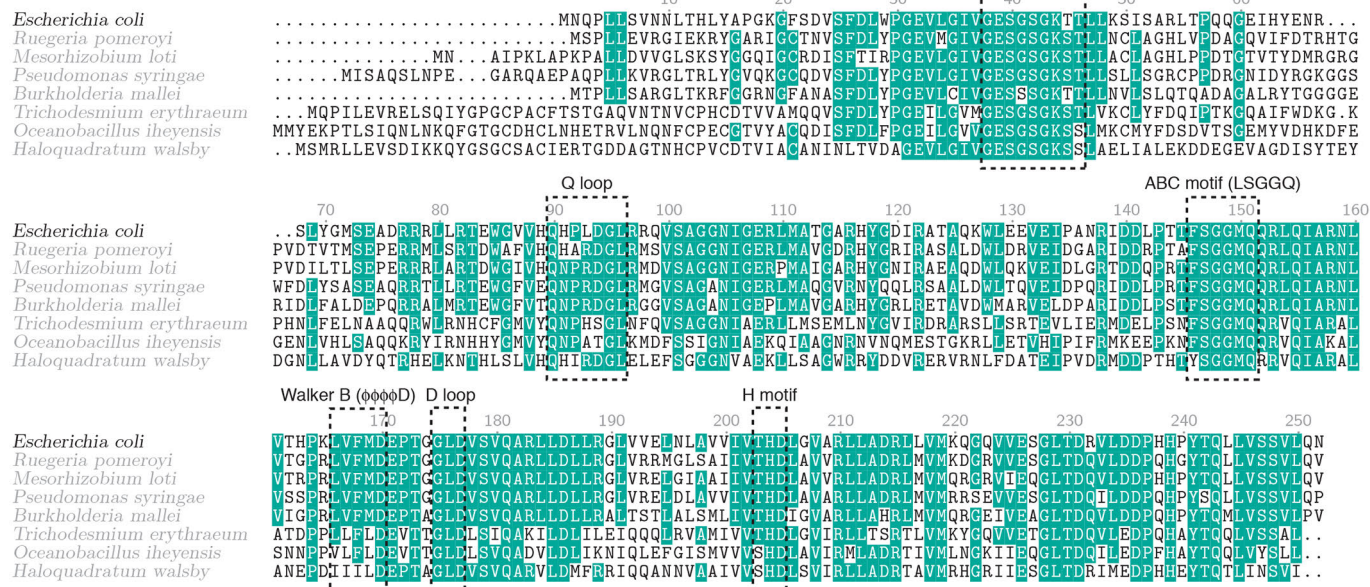
Extended Data Figure 6 | Interaction areas within the C-P lyase core complex. **a**, Dimerization interface between halves of the (PhnGHII)₂ complex. Colours as in Fig. 1. **b**, Interaction areas between the individual subunits within each dimer half. **c**, Two perpendicular views of the C-P lyase core complex shown in surface representation with colours as in Fig. 1c. **d**, Left,

overview of the surface conservation of the C-P lyase core complex shown as a colour gradient from teal (variable) to burgundy (conserved) as indicated. Right, conservation at the interaction interfaces between the individual subunits of the complex.



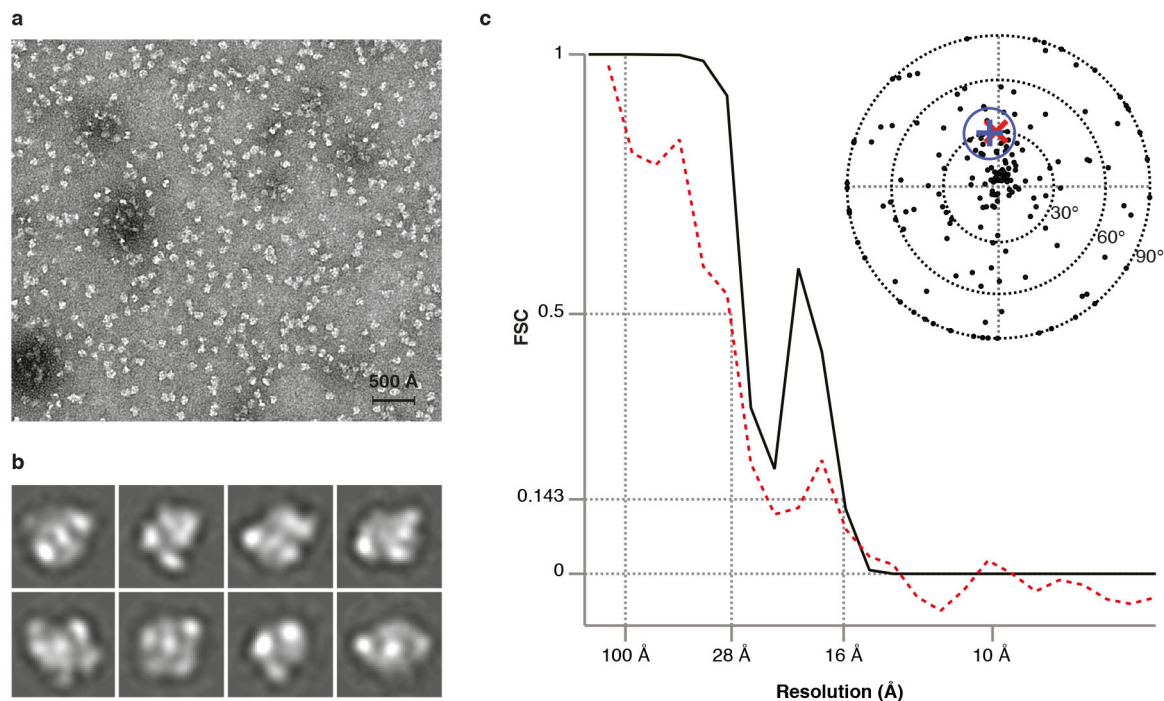
Extended Data Figure 7 | In vivo complementation of *E. coli* $\Delta phnHIJKLMNOP$. *E. coli* strain BW16711 ($\Delta phnHIJKLMNOP$) complemented with a plasmid-borne copy of either wild-type (Wt) *phnGHJKLMNOP* or variants thereof, including PhnJ(C272A), PhnJ(H108A), PhnI(H328A), or the PhnI(H328A;H333A) double variant. Growth is

monitored on minimal plates with either no phosphorus source (a), 2-aminoethyl phosphonate (2-AEPn) (b), methyl phosphonate (MPn) (c), or phosphate ion (d). The data shown are representative of three independent experiments.



Extended Data Figure 8 | PhnK sequence alignment. Alignment of the protein sequence of *E. coli* PhnK with homologous proteins from a wide range of microorganisms. Conserved residues are shown on a teal background, and

residues mentioned in the main text and the location of the ABC cassette consensus motifs (Walker A, Q motif, ABC motif, Walker B, D loop, and H motif) are indicated.



Extended Data Figure 9 | Negative stain electron microscopy of PhnGHIIJK.

a, Raw micrograph representative of 100 images collected. Scale bar is 500 Å.

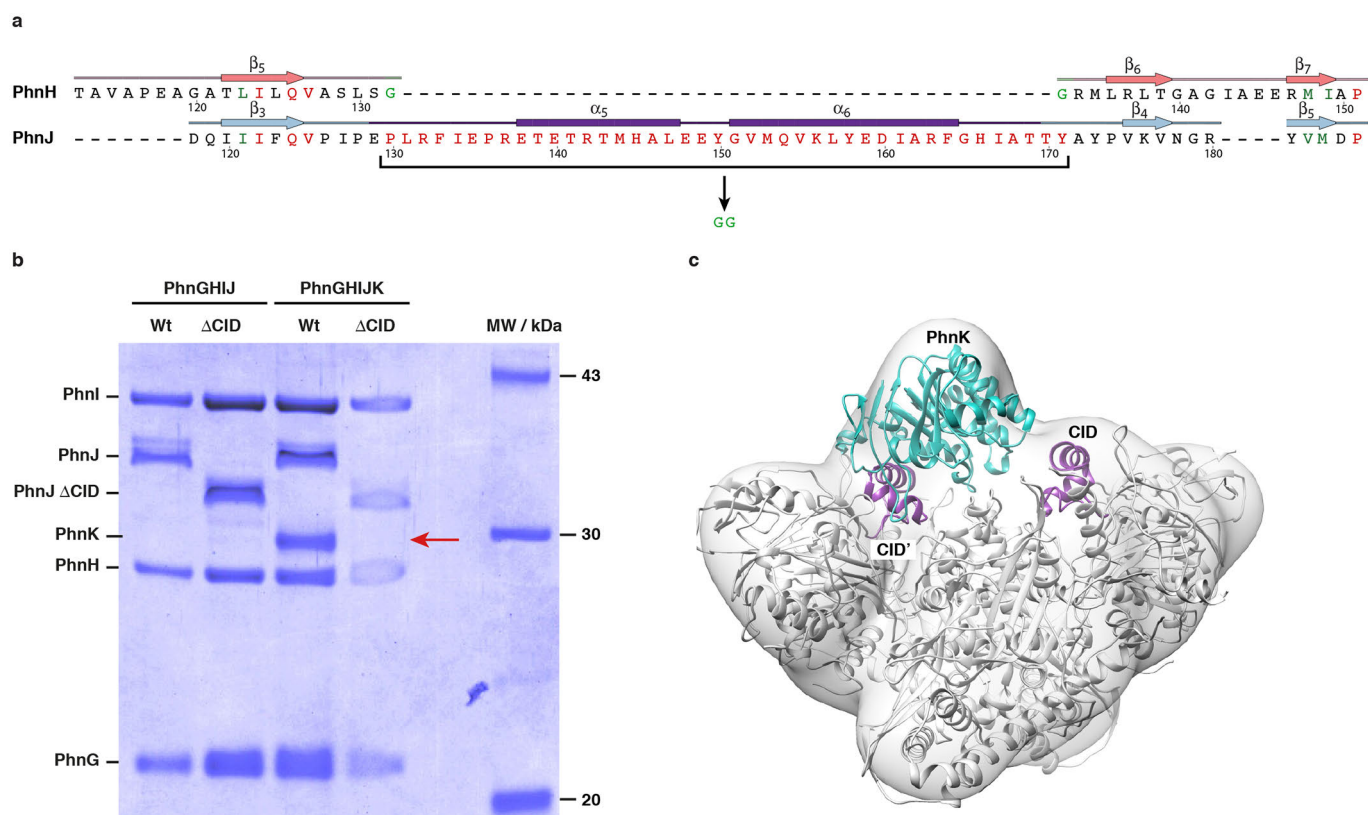
b, Selection of 2D reference-free class averages from a total of 300 classes

showing the particle in various orientations. Each class is 172 Å wide. **c**, Fourier-

shell correlation (FSC) of the final electron microscopy density map as a

function of resolution (black line) with FSC = 0.143 at 16 Å. The red line shows

the correlation between the crystal structure and the electron microscopy density, which has FSC = 0.5 at 28 Å. The inset figure shows an equal area projection plot for the electron microscopy tilt pair validation data set. The circle is an approximation of a 99% confidence interval that contains the representative direction (blue plus), includes the true tilt direction (red cross), and excludes the untitled direction (origin).



Extended Data Figure 10 | Purification of the PhnGHIJ(ΔCID) and PhnGHIJ(ΔCID)K complexes. **a**, Sequence alignment of the CID region of PhnJ with the corresponding part of the structural domain of PhnH, where strands 5 and 6 of the β -sheet are connected by a short Gly–Gly turn (green residues). In PhnJ(ΔCID), the CID domain spanning residues 130–171 (red residues) are replaced by a similar dipeptide turn, which should maintain the overall domain fold. **b**, SDS–PAGE gel showing purified C–P lyase complexes

(PhnGHIJ and PhnGHIJK) both with (Wt) and without (ΔCID) the CID on PhnJ. PhnK is missing from the complex purified from the PhnGHIJ(ΔCID)K construct (red arrow). The data shown are representative of three independent purifications. **c**, Overview of the C–P lyase core complex structure docked in the PhnGHIJK electron microscopy density with the location of the PhnJ CID domains (purple) and PhnK (cyan cartoon) indicated.

Extended Data Table 1 | Data collection, phasing and refinement statistics

	Native	Ta ₆ Br ₁₂ derivative
Data collection		
Space group	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁
Cell dimensions		
a, b, c (Å)	95.5, 133.7, 176.7	95.8, 143.1, 178.7
α, β, γ (°)	90.0, 90.0, 90.0	90.0, 90.0, 90.0
Resolution	58.91 - 1.70 Å (1.74- 1.70 Å) [*]	48.24 - 3.50 (3.60 - 3.50 Å)
Wavelength	1.00004 Å	1.2552 Å
Unique reflections	246,950 (17,488)	59,846 (4,847)
R-meas (%)	7.0 (114.7)	15.9 (48.0)
CC _{1/2}	99.9 (51.6)	99.7 (96.4)
I/ σ _I	17.7 (1.7)	17.0 (7.7)
Completeness (%)	99.7 (99.1)	100.0 (100.0)
Redundancy	5.6 (5.5)	26.6 (26.3)
Refinement		
Resolution (Å)	58.4 - 1.7	
No. of reflections	246,797	
R_{work} / R_{free} (%)	14.9 (17.6)	
No. of atoms		
Protein (non-hydrogen)	30,097 (15,203)	
SO ₄ ²⁻ / Zn ²⁺	24	
Water	1,792	
B-factors (Å ²)		
Protein	29.87	
SO ₄ ²⁻	40.28	
Zn ²⁺	25.97	
Water	21.15	
R.m.s deviations		
Bond lengths (Å)	0.01	
Bond angles (°)	1.17	
Ramachandran statistics [†]		
Favoured (%)	97.5	
Allowed (%)	2.0	
Outliers (%)	0.5	

* Highest resolution shell is shown in parentheses, except where otherwise stated.

† Statistics from MolProbity via Phenix³⁸.

Conventional superconductivity at 203 kelvin at high pressures in the sulfur hydride system

A. P. Drozdov^{1*}, M. I. Erements^{1*}, I. A. Troyan¹, V. Ksenofontov² & S. I. Shylin²

A superconductor is a material that can conduct electricity without resistance below a superconducting transition temperature, T_c . The highest T_c that has been achieved to date is in the copper oxide system¹: 133 kelvin at ambient pressure² and 164 kelvin at high pressures³. As the nature of superconductivity in these materials is still not fully understood (they are not conventional superconductors), the prospects for achieving still higher transition temperatures by this route are not clear. In contrast, the Bardeen–Cooper–Schrieffer theory of conventional superconductivity gives a guide for achieving high T_c with no theoretical upper bound—all that is needed is a favourable combination of high-frequency phonons, strong electron–phonon coupling, and a high density of states⁴. These conditions can in principle be fulfilled for metallic hydrogen and covalent compounds dominated by hydrogen^{5,6}, as hydrogen atoms provide the necessary high-frequency phonon modes as well as the strong electron–phonon coupling. Numerous calculations support this idea and have predicted transition temperatures in the range 50–235 kelvin for many hydrides⁷, but only a moderate T_c of 17 kelvin has been observed experimentally⁸. Here we investigate sulfur hydride⁹, where a T_c of 80 kelvin has been predicted¹⁰. We find that this system transforms to a metal at a pressure of approximately 90 gigapascals. On cooling, we see signatures of superconductivity: a sharp drop of the resistivity to zero and a decrease of the transition temperature with magnetic field, with magnetic susceptibility measurements confirming a T_c of 203 kelvin. Moreover, a pronounced isotope shift of T_c in sulfur deuteride is suggestive of an electron–phonon mechanism of superconductivity that is consistent with the Bardeen–Cooper–Schrieffer scenario. We argue that the phase responsible for high- T_c superconductivity in this system is likely to be H_3S , formed from H_2S by decomposition under pressure. These findings raise hope for the prospects for achieving room-temperature superconductivity in other hydrogen-based materials.

A search for high- (room)-temperature conventional superconductivity is likely to be fruitful, as the Bardeen–Cooper–Schrieffer (BCS) theory in the Eliashberg formulation puts no apparent limits on T_c . Materials with light elements are especially favourable as they provide high frequencies in the phonon spectrum. Indeed, many superconductive materials have been found in this way, but only a moderately high $T_c = 39$ K has been found in this search (in MgB_2 ; ref. 11).

Ashcroft⁵ turned attention to hydrogen, which has very high vibrational frequencies due to the light hydrogen atom and provides a strong electron–phonon interaction. Further calculations showed that metallic hydrogen should be a superconductor with a very high T_c of about 100–240 K for molecular hydrogen, and of 300–350 K in the atomic phase at 500 GPa (ref. 12). However, superconductivity in pure hydrogen has not yet been found, even though a conductive and probably semimetallic state of hydrogen has been recently produced¹³. Hydrogen-dominated materials such as covalent hydrides SiH_4 , SnH_4 , and so on might also be good candidates for showing high- T_c

superconductivity⁶. Similarly to pure hydrogen, they have high Debye temperatures. Moreover, heavier elements might be beneficial as they contribute to the low frequencies that enhance electron–phonon coupling. Importantly, lower pressures are required to metallize hydrides in comparison to pure hydrogen. Ashcroft's general idea was supported in numerous calculations^{7,10} predicting high values of T_c for many hydrides. So far only a low T_c (~ 17 K) has been observed experimentally⁸.

For the present study we selected H_2S , because it is relatively easy to handle and is predicted to transform to a metal and a superconductor at a low pressure $P \approx 100$ GPa with a high $T_c \approx 80$ K (ref. 10). Experimentally, H_2S is known as a typical molecular compound with a rich phase diagram¹⁴. At about 96 GPa, hydrogen sulphide transforms to a metal¹⁵. The transformation is complicated by the partial dissociation of H_2S and the appearance of elemental sulfur at $P > 27$ GPa at room temperature, and at higher pressures at lower temperatures¹⁴. Therefore, the metallization of hydrogen sulphide can be explained by elemental sulfur, which is known to become metallic above 95 GPa (ref. 16). No experimental studies of hydrogen sulphide are known above 100 GPa.

In a typical experiment, we performed loading and the initial pressure increase at temperatures of ~ 200 K; this is essential for obtaining a good sample (Methods). The Raman spectra of H_2S and D_2S were measured as the pressure was increased, and were in general agreement with the literature data^{17,18} (Extended Data Fig. 1). The sample starts to conduct at $P \approx 50$ GPa. At this pressure it is a semiconductor, as shown by the temperature dependence of the resistance and pronounced photoconductivity. At 90–100 GPa the resistance drops further, and the temperature dependence becomes metallic. No photoconductive response is observed in this state. It is a poor metal—its resistivity at ~ 100 K is $\rho \approx 3 \times 10^{-5}$ ohm m at 110 GPa and $\rho \approx 3 \times 10^{-7}$ ohm m at ~ 200 GPa.

During the cooling of the metal at pressures of about 100 GPa (Fig. 1a) the resistance abruptly drops by three to four orders of magnitude, indicating a transition to the superconducting state. At the next increase of pressure at low temperatures of $T < 100$ K, T_c steadily increases with pressure. However, at pressures of > 160 GPa, T_c increases sharply (Fig. 1b). As higher temperatures of 150–250 K were involved in this pressure range, we supposed that the increase of T_c and the decrease of sample resistance during warming (Fig. 1a) could indicate a possible kinetic-controlled phase transformation. Therefore in further experiments, after loading and after the initial pressure increase at 200 K, we annealed all samples by heating them to room temperature (or above) at pressures of $> \sim 150$ GPa (Fig. 2a, see also Extended Data Fig. 2). This allowed us to obtain stable results, to compare different isotopes, to obtain the dependence of T_c on pressure and magnetic field, and to prove the existence of superconductivity in our samples as follows. (We note that additional information on experimental conditions are given in the appropriate figure legends.)

¹Max-Planck-Institut für Chemie, Hahn-Meitner-Weg 1, 55128 Mainz, Germany. ²Institut für Anorganische Chemie und Analytische Chemie, Johannes Gutenberg-Universität Mainz, Staudingerweg 9, 55099 Mainz, Germany.

*These authors contributed equally to this work.

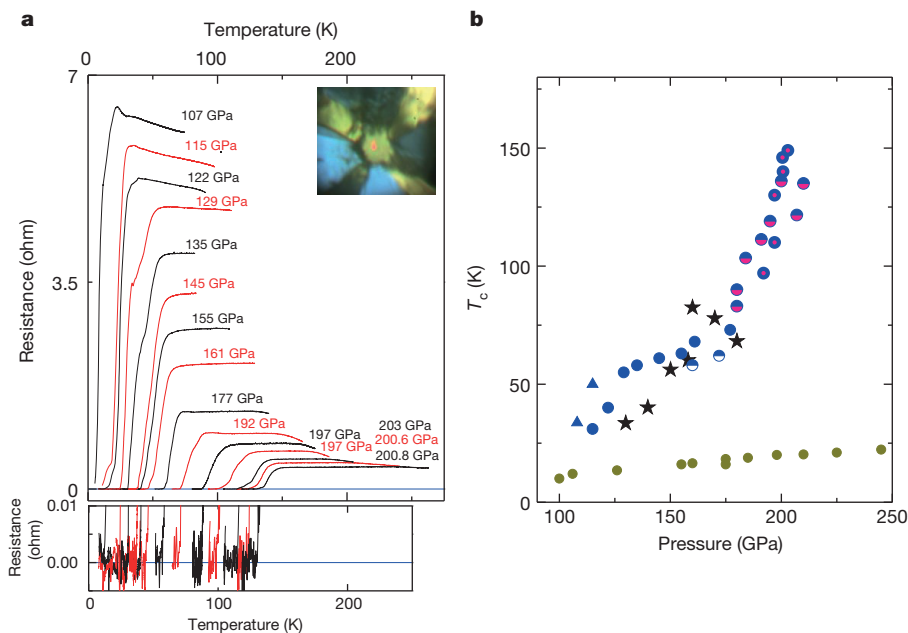


Figure 1 | Temperature dependence of the resistance of sulfur hydride measured at different pressures, and the pressure dependence of T_c . **a**, Main panel, temperature dependence of the resistance (R) of sulfur hydride at different pressures. The pressure values are indicated near the corresponding plots. At first, the sample was loaded at $T \approx 200$ K and the pressure was increased to ~ 100 GPa; the sample was then cooled down to 4 K. After warming to ~ 100 K, pressure was further increased. Plots at pressures < 135 GPa have been scaled (reduced) as follows—105 GPa, by 10 times; 115 GPa and 122 GPa, by 5 times; and 129 GPa by 2 times—for easier comparison with the higher pressure steps. The resistance was measured with a current of 10 μ A. Bottom panel, the resistance plots near zero. The resistance was measured with four electrodes deposited on a diamond anvil that touched the sample (top panel inset). The diameters of the samples were ~ 25 μ m and the thickness was

~ 1 μ m. **b**, Blue round points represent values of T_c determined from **a**. Other blue points (triangles and half circles) were obtained in similar runs. Measurements at $P > \sim 160$ GPa revealed a sharp increase of T_c . In this pressure range the $R(T)$ measurements were performed over a larger temperature range up to 260 K, the corresponding experimental points for two samples are indicated by adding a pink colour to half circles and a centred dot to filled circles. These points probably reflect a transient state for these particular P/T conditions. Further annealing of the sample at room temperature would require stabilizing the sample (Fig. 2a). Black stars are calculations from ref. 10. Dark yellow points are T_c values of pure sulfur obtained with the same four-probe electrical measurement method. They are consistent with literature data³⁰ (susceptibility measurements) but have higher values at $P > 200$ GPa.

(1) There is a sharp drop in resistivity with cooling, indicating a phase transformation. The measured minimum resistance is at least as low, $\sim 10^{-11}$ ohm m—about two orders of magnitude less than for pure copper (Fig. 1, Extended Data Fig. 3e) measured at the same temperature¹⁹. (2) A strong isotope effect is observed: T_c shifts to lower

temperatures for sulfur deuteride, indicating phonon-assisted superconductivity (Fig. 2b, c). The BCS theory gives the dependence of T_c on atomic mass m as $T_c \propto m^{-\alpha}$, where $\alpha \approx 0.5$. Comparison of T_c values in the pressure range $P > 170$ GPa (Fig. 2c) gives $\alpha \approx 0.3$. (3) T_c shifts to lower temperatures with available magnetic field (B) up to 7 T

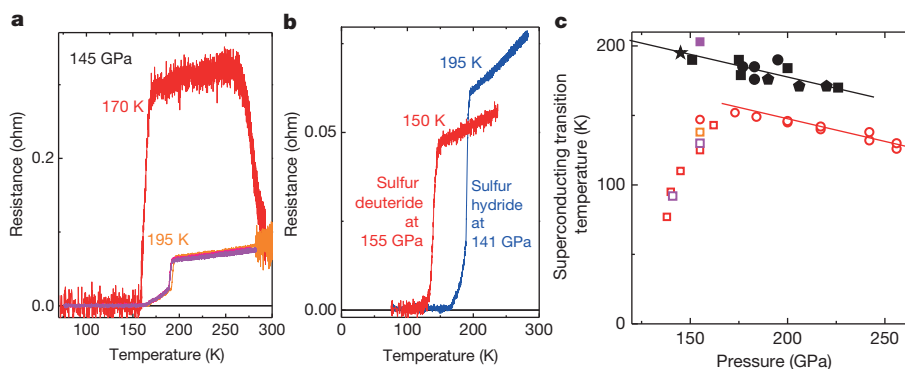


Figure 2 | Pressure and temperature effects on T_c of sulfur hydride and sulfur deuteride. **a**, Changes of resistance and T_c of sulfur hydride with temperature at constant pressure—the annealing process. The sample was pressurized to 145 GPa at 220 K and then cooled to 100 K. It was then slowly warmed at ~ 1 K min^{-1} ; $T_c = 170$ K was determined. At temperatures above ~ 250 K the resistance dropped sharply, and during the next temperature run T_c increased to ~ 195 K. This T_c remained nearly the same for the next two runs. (We note that the only point for sulfur deuteride presented in ref. 9 was determined without sample annealing, and T_c would increase after annealing at room temperature.) **b**, Typical superconductive steps for sulfur hydride

(blue trace) and sulfur deuteride (red trace). The data were acquired during slow warming over a time of several hours. T_c is defined here as the sharp kink in the transition to normal metallic behaviour. These curves were obtained after annealing at room temperature as shown in **a**. **c**, Dependence of T_c on pressure; data on annealed samples are presented. Open coloured points refer to sulfur deuteride, and filled points to sulfur hydride. Data shown as the magenta point were obtained in magnetic susceptibility measurements (Fig. 4a). The lines indicate that the plots are parallel at pressures above ~ 170 GPa (the isotope shift is constant) but strongly deviate at lower pressures.

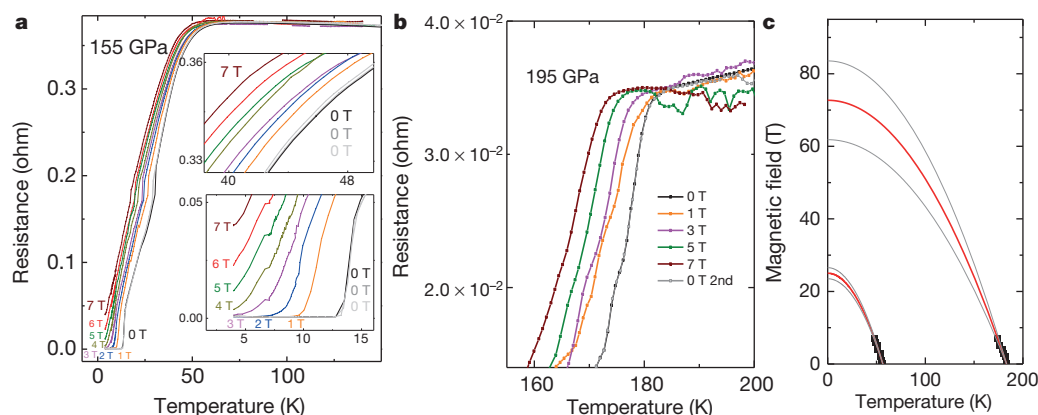


Figure 3 | Temperature dependence of the resistance of sulfur hydride in different magnetic fields. **a**, The shift of the ~ 60 K superconducting transition in magnetic fields of 0–7 T (colour coded). The upper and lower parts of the transition are shown enlarged in the insets (axes as in main panel). The temperature dependence of the resistance without an applied magnetic field was measured three times: before applying the field, after applying 1, 3, 5, 7 T and finally after applying 2, 4, 6 T (black, grey and dark grey colours). **b**, The

same measurements but for the 185 K superconducting transition. **c**, The temperature dependence of the critical magnetic field strengths of sulfur hydride. T_c (black points deduced from **a**, **b**) are plotted for the corresponding magnetic fields. To estimate the critical magnetic field H_c , the plots were extrapolated to high magnetic fields using the formula $H_c(T) = H_{c0}(1 - (T/T_c)^2)$. The extrapolation has been done with 95% confidence (band shown as grey lines).

(Fig. 3). Much higher fields are required to destroy the superconductivity: extrapolation of $T_c(B)$ gives an estimate of a critical magnetic field as high as 70 T (Fig. 3). (4) Finally, in magnetic susceptibility measurements (Fig. 4) a sharp transition from the diamagnetic to the paramagnetic state (Fig. 4a) was observed for zero-field-cooled (ZFC) material. The onset temperature of the superconducting state $T_{\text{onset}} = 203(1)$ K, and the width of the superconducting transition is nearly the same as in electrical measurements (Fig. 4a). Magnetization measurements $M(H)$, where H is magnetic field, at different

temperatures (Fig. 4c) revealed a pronounced hysteresis indicating type II superconductivity with the first critical field $H_{c1} \approx 30$ mT. The magnetization decreases sharply at temperatures above 200 K showing the onset of superconductivity at 203.5 K, in agreement with the susceptibility measurements (Fig. 4a). A list of key properties of the new superconductor is given in Methods.

We have presented purely experimental evidence of superconductivity in sulfur hydride. However the particular compound responsible for the high T_c is not obvious. The superconductivity measured in the

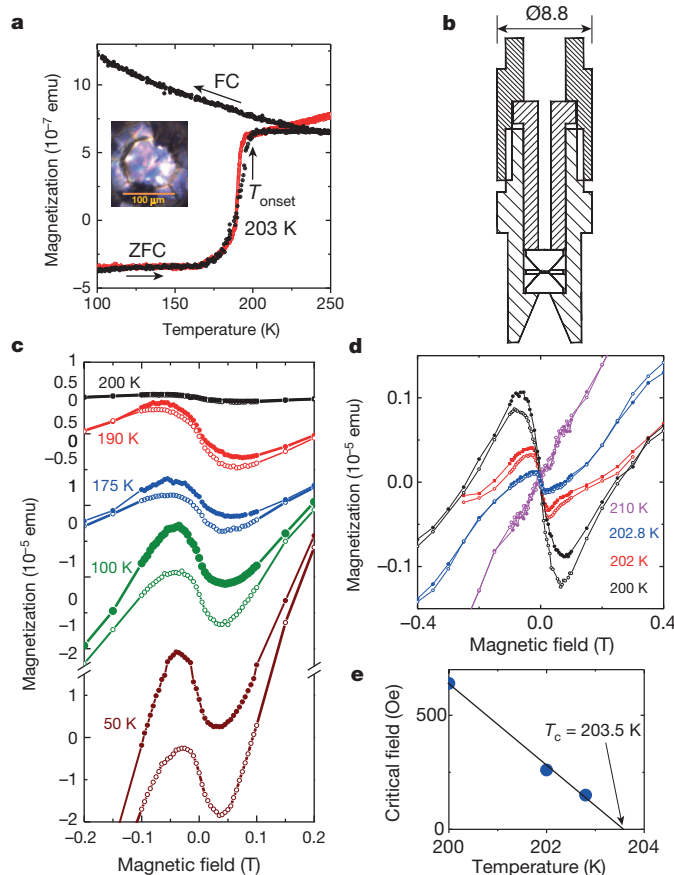


Figure 4 | Magnetization measurements. **a**, Temperature dependence of the magnetization of sulfur hydride at a pressure of 155 GPa in zero-field cooled (ZFC) and 20 Oe field cooled (FC) modes (black circles). The onset temperature is $T_{\text{onset}} = 203(1)$ K. For comparison, the superconducting step obtained for sulfur hydride from electrical measurements at 145 GPa is shown by red circles. Resistivity data ($T_{\text{onset}} = 195$ K) were scaled and moved vertically to compare with the magnetization data. Inset, optical micrograph of a sulfur hydride sample at 155 GPa in a CaSO_4 gasket (scale bar 100 μm). The high $T_{\text{onset}} = 203$ K measured from the susceptibility can be explained by a significant input to the signal from the periphery of the sample which expanded beyond the culet where pressure is smaller than in the culet centre (T_c increases with decreasing pressure (Fig. 2b)). **b**, Non-magnetic diamond anvil cell (DAC) of diameter 8.8 mm. **c**, Magnetization measurements $M(H)$ of sulfur hydride at a pressure of 155 GPa at different temperatures (given as curve labels). The magnetization curves show hysteresis, indicating a type II superconductor. The magnetization curves are however distorted by obvious paramagnetic input (which is also observed in other superconductors³¹). In our case, the paramagnetic signal is probably from the DAC, but further study of the origin of this input is required. The paramagnetic background increases when temperature is decreased. The minima of the magnetization curves (~ 35 mT) are the result of the diamagnetic input from superconductivity and the paramagnetic background. The first critical field $H_{c1} \approx 30$ mT can be roughly estimated as the point where magnetization deviates from linear behaviour. At higher fields, magnetization increases due to the penetration of magnetic vortices. As the sign of the field change reverses, the magnetic flux in the Shubnikov phase remains trapped and therefore the back run (that is, with decreasing field) is irreversible—the returning branch of the magnetic cycle (shown by filled points) runs above the direct one. Hysteretic behaviour of the magnetization becomes more clearly visible as the temperature decreases. **d**, At high temperatures $T > 200$ K, the magnetization decreases sharply. **e**, Extrapolation of the pronounced minima at the magnetization curves to higher temperatures gives the onset of superconductivity at $T = 203.5$ K.

low-temperature runs (Fig. 1) possibly relates to H_2S , as it is generally consistent with calculations¹⁰ for H_2S : both the value of $T_c \approx 80$ K and its pressure behaviour. However superconductivity with $T_c \approx 200$ K (Fig. 2) does not follow from these calculations. We suppose that it relates to the decomposition of H_2S , as high temperatures are required to reach the high T_c (Fig. 2b). Precipitation of elemental sulfur on decomposition could be expected (which is well known at low pressures of $P < 100$ GPa; ref. 14); however the superconducting transition in elemental sulfur occurs at significantly lower temperatures (Fig. 1b). Another expected product of decomposition of H_2S is hydrogen. However, the strong characteristic vibrational stretching mode from the H_2 molecule was never observed in our Raman spectra (nor was it observed in ref. 14). Therefore we suppose that the dissociation of H_2S is different and involves the creation of higher hydrides, such as $3\text{H}_2\text{S} \rightarrow \text{H}_6\text{S} + 2\text{S}$ or $2\text{H}_2\text{S} \rightarrow \text{H}_4\text{S} + \text{S}$. It is natural to expect these reactions, as sulfur can be not only divalent, but also exhibits higher valencies. In fact, calculations¹⁰ indirectly support this hypothesis, as the dissociation $\text{H}_2\text{S} \rightarrow \text{H}_2 + \text{S}$ was shown to be energetically very unfavourable. We found further theoretical support in ref. 20. In that work, the van der Waals compound²¹ $(\text{H}_2\text{S})_2\text{H}_2$ was considered, and it was shown that at pressures above 180 GPa it forms an $Im\text{-}3m$ structure with H_3S stoichiometry. The predicted $T_c \approx 190$ K and its pressure dependences are close to our experimental values (Fig. 2c). Our hypothesis of the transformation of H_2S to higher hydrides (in the H_3S stoichiometry each S atom is surrounded by 6 hydrogen atoms) is strongly supported by further calculations^{22,23}. All the numerous works based on the $Im\text{-}3m$ structure^{23–27} are consistent in their prediction of $T_c > \sim 200$ K, which decreases with pressure. The hydrogen sublattice gives the main contribution to superconductivity^{20,25,26}. Inclusion of zero point vibrations and anharmonicity in the calculations²⁴ corrected the calculated T_c to ~ 190 K, and the isotope coefficient from $\alpha = 0.5$ to $\alpha = 0.35$ —both in agreement with the present work.

The highest T_c of 203 K that we report here has been achieved most probably in H_3S having the $Im\text{-}3m$ structure. It is a good metal; interestingly, there is also strong covalent bonding between H and S atoms in this compound²⁰. This is in agreement with the general assumption (see for instance ref. 28) that a metal with high T_c should have strong covalent bonding (as is realized in MgB_2 ; ref. 29) together with high-frequency modes in the phonon spectrum. This particular combination of bonding type and phonon spectrum would probably provide a good criterion when searching for the materials with high T_c at ambient pressure that are required for applications. There are many hydrogen-containing materials with strong covalent bonding (such as organics) but typically they are insulators. In principle, they could be tuned to a metallic state by doping or gating. Modern methods of structure prediction could facilitate exploration for the desired materials.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 25 June; accepted 22 July 2015.

Published online 17 August 2015.

1. Bednorz, J. G. & Mueller, K. A. Possible high T_c superconductivity in the Ba-La-Cu-O system. *Z. Phys. B* **64**, 189–193 (1986).
2. Schilling, A., Cantoni, M., Guo, J. D. & Ott, H. R. Superconductivity above 130 K in the Hg-Ba-Ca-Cu-O system. *Nature* **363**, 56–58 (1993).
3. Gao, L. *et al.* Superconductivity up to 164 K in $\text{HgBa}_2\text{Ca}_{m-1}\text{Cu}_m\text{O}_{2m+2+\delta}$ ($m=1, 2$, and 3) under quasihydrostatic pressures. *Phys. Rev. B* **50**, 4260–4263 (1994).
4. Ginzburg, V. L. Once again about high-temperature superconductivity. *Contemp. Phys.* **33**, 15–23 (1992).
5. Ashcroft, N. W. Metallic hydrogen: A high-temperature superconductor? *Phys. Rev. Lett.* **21**, 1748–1750 (1968).
6. Ashcroft, N. W. Hydrogen dominant metallic alloys: high temperature superconductors? *Phys. Rev. Lett.* **92**, 187002 (2004).

7. Wang, Y. & Ma, Y. Perspective: Crystal structure prediction at high pressures. *J. Chem. Phys.* **140**, 040901 (2014).
8. Erements, M. I., Trojan, I. A., Medvedev, S. A., Tse, J. S. & Yao, Y. Superconductivity in hydrogen dominant materials: silane. *Science* **319**, 1506–1509 (2008).
9. Drozdov, A. P., Erements, M. I. & Troyan, I. A. Conventional superconductivity at 190 K at high pressures. Preprint at <http://arXiv.org/abs/1412.0460> (2014).
10. Li, Y., Hao, J., Li, Y. & Ma, Y. The metallization and superconductivity of dense hydrogen sulfide. *J. Chem. Phys.* **140**, 174712 (2014).
11. Nagamatsu, J., Nakagawa, N., Muranaka, T., Zenitani, Y. & Akimitsu, J. Superconductivity at 39 K in magnesium diboride. *Nature* **410**, 63–64 (2001).
12. McMahon, J. M., Morales, M. A., Pierleoni, C. & Ceperley, D. M. The properties of hydrogen and helium under extreme conditions. *Rev. Mod. Phys.* **84**, 1607–1653 (2012).
13. Erements, M. I. & Troyan, I. A. Conductive dense hydrogen. *Nature Mater.* **10**, 927–931 (2011).
14. Fujihisa, H. *et al.* Molecular dissociation and two low-temperature high-pressure phases of H_2S . *Phys. Rev. B* **69**, 214102 (2004).
15. Sakashita, M. *et al.* Pressure-induced molecular dissociation and metallization in hydrogen-bonded H_2S solid. *Phys. Rev. Lett.* **79**, 1082–1085 (1997).
16. Kometani, S., Erements, M., Shimizu, K., Kobayashi, M. & Amaya, K. Observation of pressure-induced superconductivity of sulfur. *J. Phys. Soc. Jpn.* **66**, 2564–2565 (1997).
17. Shimizu, H. *et al.* Pressure-temperature phase diagram of solid hydrogen sulfide determined by Raman spectroscopy. *Phys. Rev. B* **51**, 9391–9394 (1995).
18. Shimizu, H., Murashima, H. & Sasaki, S. High-pressure Raman study of solid deuterium sulfide up to 17 GPa. *J. Chem. Phys.* **97**, 7137–7139 (1992).
19. Matula, R. A. Electrical resistivity of copper, gold, palladium, and silver. *J. Phys. Chem. Ref.* **8**, 1147–1298 (1979).
20. Duan, D. *et al.* Pressure-induced metallization of dense $(\text{H}_2\text{S})_2\text{H}_2$ with high- T_c superconductivity. *Sci. Rep.* **4**, 6968 (2014).
21. Strobel, T. A., Ganesh, P., Somayazulu, M., Kent, P. R. C. & Hemley, R. J. Novel cooperative interactions and structural ordering in H_2S – H_2 . *Phys. Rev. Lett.* **107**, 255503 (2011).
22. Duan, D. *et al.* Pressure-induced decomposition of solid hydrogen sulfide. *Phys. Rev. B* **91**, 180502(R) (2015).
23. Bernstein, N., Hellberg, C. S., Johannes, M. D., Mazin, I. I. & Mehl, M. J. What superconductors in sulfur hydrides under pressure, and why. *Phys. Rev. B* **91**, 060511(R) (2015).
24. Errea, I. *et al.* Hydrogen sulfide at high pressure: a strongly-anharmonic phonon-mediated superconductor. *Phys. Rev. Lett.* **114**, 157004 (2015).
25. Flores-Livas, J. A., Sanna, A. & Gross, E. K. U. High temperature superconductivity in sulfur and selenium hydrides at high pressure. Preprint at <http://arXiv.org/abs/1501.06336v1> (2015).
26. Papaconstantopoulos, D. A., Klein, B. M., Mehl, M. J. & Pickett, W. E. Cubic H_3S around 200 GPa: an atomic hydrogen superconductor stabilized by sulfur. *Phys. Rev. B* **91**, 184511 (2015).
27. Akashi, R., Kawamura, M., Tsuneyuki, S., Nomura, Y. & Arita, R. Fully non-empirical study on superconductivity in compressed sulfur hydrides. Preprint at <http://arXiv.org/abs/1502.00936v1> (2015).
28. Cohen, M. L. in *BCS: 50 years* (eds Cooper, L. N. & Feldman, D.) 375–389 (World Scientific, 2011).
29. An, J. M. & Pickett, W. E. Superconductivity of MgB_2 : covalent bonds driven metallic. *Phys. Rev. Lett.* **86**, 4366–4369 (2001).
30. Gregoryanz, E. *et al.* Superconductivity in the chalcogens up to multimegabar pressures. *Phys. Rev. B* **65**, 064504 (2002).
31. Senoussi, S., Sastry, P., Yakhmi, J. V. & Campbell, I. Magnetic hysteresis of superconducting $\text{GdBa}_2\text{Cu}_3\text{O}_7$ down to 1.8 K. *J. Phys.* **49**, 2163–2164 (1988).

Acknowledgements Support provided by the European Research Council under the 2010 Advanced Grant 267777 is acknowledged. We appreciate help provided in MPI Chemie by U. Pöschl. We thank P. Alireza and G. Lonzarich for help with samples of CuTi ; J. Kamarad, S. Toser and C. Q. Jin for sharing their experience on SQUID measurements; K. Shimizu and his group for cooperation; P. Chu and his group for many discussions and collaboration, and L. Pietronero, M. Calandra and T. Timusk for discussions. V.K. and S.I.S. acknowledge the DFG (Priority Program No. 1458) for support. M.I.E. thanks H. Musshof and R. Wittkowski for precision machining of the DACs.

Author Contributions A.P.D. performed the most of the experiments and contributed to the data interpretation and writing the manuscript. M.I.E. designed the study, wrote the major part of the manuscript, developed the DAC for SQUID measurements, and participated in the experiments. I.A.T. participated in experiments. V.K. and S.I.S. performed the magnetic susceptibility measurements and contributed to writing the manuscript. M.I.E. and A.P.D. contributed equally to this paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.I.E. (m.erements@mpic.de).

METHODS

Experimental procedure. For electrical measurements we used diamond anvils (DACs) with anvils of the following shape: tip diameter of 200–300 μm bevelled at $7\text{--}8^\circ$ to a culet of 40–80 μm . An insulating gasket is required to separate the metallic gasket from the electrodes. It was prepared in the following way (Extended Data Fig. 3). First a metallic gasket of T301 stainless steel (or Re) 250 μm thick was indented with about 17–20 GPa pressure. Then the bottom of the imprint of diameter $\sim 200\text{ }\mu\text{m}$ was drilled out, and a powder insulating material was put in the imprint and pressed between the anvils to form a layer. The insulating layer was made of either Teflon, NaCl or CaSO_4 as these materials do not react with H_2S . The layer was pressed to obtain a thickness in the centre of $\sim 3\text{--}5\text{ }\mu\text{m}$ to provide stable clamping. A larger thickness leads to instability in the sample—it shifts or escapes under pressure—while with a thinner gasket it is difficult to reach high pressures. A hole of diameter $\sim 10\text{--}30\text{ }\mu\text{m}$ was then drilled in the insulating layer. Four Ti electrodes were sputtered on the diamond anvil. The electrodes were capped with Au to prevent oxidation of the Ti. (To check a possible contribution of the diamond surface to the conductivity, we prepared a different configuration of electrodes for a once-only experiment: two electrodes were sputtered on one anvil and another two on another anvil, similar to ref. 13). After preparation of the electrodes the gasket was put back on the anvil and the DAC was assembled so that the separation between the anvils was about 20–100 μm (measured by interference fringes). The DAC was placed into a cryostat and cooled down to $\sim 200\text{ K}$ (within the temperature range of liquid H_2S) and then H_2S gas was put through a capillary into a rim around the diamond anvil where it liquefied (Extended Data Fig. 4). H_2S of 99.5% and D_2S of 97% purity were been used. The filling was monitored visually (Extended Data Figs 4, 5) and the sample was identified by measuring Raman spectra. Then liquid H_2S was clamped in the gasket hole by pushing the piston of the DAC with the aid of screws outside the cryostat. The thickness of the sample can be estimated to be few micrometres, as measured from interference spectra through the clamped transparent sample. The thickness might be $\sim 1\text{ }\mu\text{m}$ if the sample expanded over the culet (Fig. 4). After the clamping, the DAC was heated to $\sim 220\text{ K}$ to evaporate the rest of the H_2S , and then the pressure was further increased at this temperature. The pressure remained stable during the cooling within $\pm 5\text{ GPa}$. The pressure was determined by a diamond edge scale at room temperature and low temperatures³². For optical measurements a Raman spectrometer was equipped with a nitrogen-cooled CCD and notch filters. The 632.8 nm line of a He–Ne laser was used to excite the Raman spectra and to determine pressure.

The low temperature loading seems to be required to prepare samples with high T_c . If H_2S was loaded at room temperature in the gas loader, for example, only sulfur was detected in Raman and X-ray scattering. Apparently in this route the sample decomposes before reaching the required high-pressure phase of H_3S . We did not explore all (P, T) paths to reach the state with high T_c . We found however that superconductivity is not observed in sample loaded at $\sim 200\text{ K}$ but heated to room temperature at low pressure $< \sim 100\text{ GPa}$.

The resistance and Raman spectra were measured during the pressurizing using the four-probe van der Pauw method (Extended Data Fig. 3) with a current of 10–10,000 μA . The temperature was reliably determined by using a slow warming rate ($\sim 1\text{ K min}^{-1}$) and allowing the DAC to equilibrate with attached thermometer. The determined T_c was well reproduced in measurements with the PPMS6000 (Physical Property Measurement System from Quantum Design)

and other set-ups. T_c was determined as the point of steepest change of resistance from the normal state (Fig. 2b).

The influence of the magnetic field on superconducting transitions has been measured with a non-magnetic DAC (diameter 25 mm) in a PPMS6000 in a 4–300 K temperature range and fields up to 7 T.

Magnetic susceptibility measurements were performed in an MPMS (Magnetic Property Measurement System) from Quantum Design. For these measurements a miniature non-magnetic cell made of Cu:Ti alloy working up to 200 GPa was designed (Fig. 4b). Samples of diameter $\sim 50\text{--}100\text{ }\mu\text{m}$ and a thickness of a few micrometres were prepared to provide a sufficient signal. Magnetic susceptibility measurements using a high-pressure cell were performed using a background subtraction feature of the MPMS software of the SQUID magnetometer (Extended Data Fig. 6).

Results. We present here some important key features of our new high- T_c sulfur hydride superconductor:

(1) The new superconductor is of type II. This fact is clearly supported by (i) a difference in temperature-dependent ZFC and FC magnetization (Fig. 4a), which is due to the Meissner effect (ZFC) and magnetic flux capture when the sample is cooled down from its normal state (FC); and (ii) the magnetic hysteresis curves (Fig. 4c, d). The magnetic hysteresis curves also have all the features of typical type II superconductors with a mixed state between H_{c1} and H_{c2} .

(2) A typical value of the coherence length ξ_{GL} in the framework of the Ginzburg–Landau theory can be estimated on the basis of the measured upper critical fields from conductivity measurements (Fig. 3c). Using the experimental estimation $60\text{ T} < H_{c2} < 80\text{ T}$ and the relation

$$\xi_{\text{GL}} = \frac{1}{2} \sqrt{\frac{\hbar}{\pi e H_{c2}}}$$

we find limits for the coherence length: $2.3\text{ nm} > \xi_{\text{GL}} > 2.0\text{ nm}$. We note that this relatively short coherence length is of the same order as, for instance, the values for superconducting $\text{YBa}_2\text{Cu}_3\text{O}_7$ (1.3 nm) and Nb_3Sn (3.5 nm).

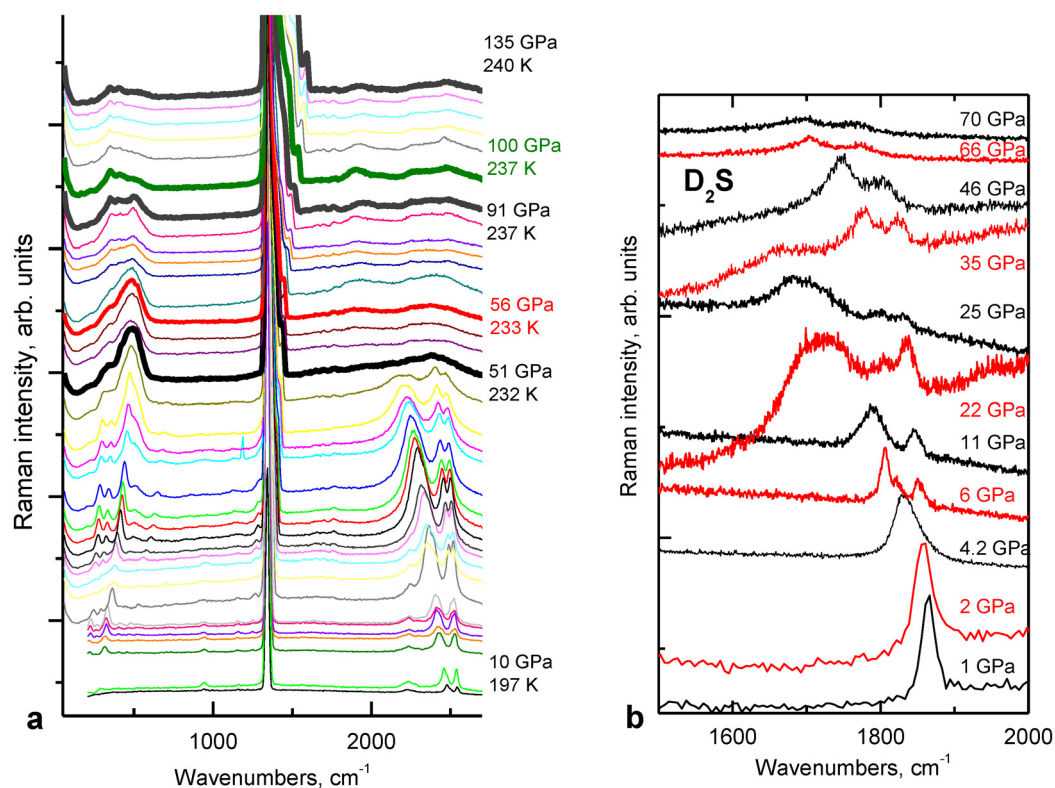
(3) The London penetration depth λ_L can be estimated from the known relation of the lower critical field H_{c1} to the upper critical field H_{c2} for a type II superconductor

$$\frac{H_{c1}}{H_{c2}} \approx \frac{\ln \kappa}{2\sqrt{2}\kappa^2}$$

in the limit $\kappa \gg 1$ of the Ginzburg–Landau parameter $\kappa = \frac{\lambda_L}{\xi_{\text{GL}}}$. Considering the experimental value of the first critical field of $3 \times 10^{-2}\text{ T}$ (Fig. 4c) and the above-mentioned relation $60\text{ T} < H_{c2} < 80\text{ T}$, we can obtain the following estimate for the London penetration depth: $\lambda_L \approx 125\text{ nm}$.

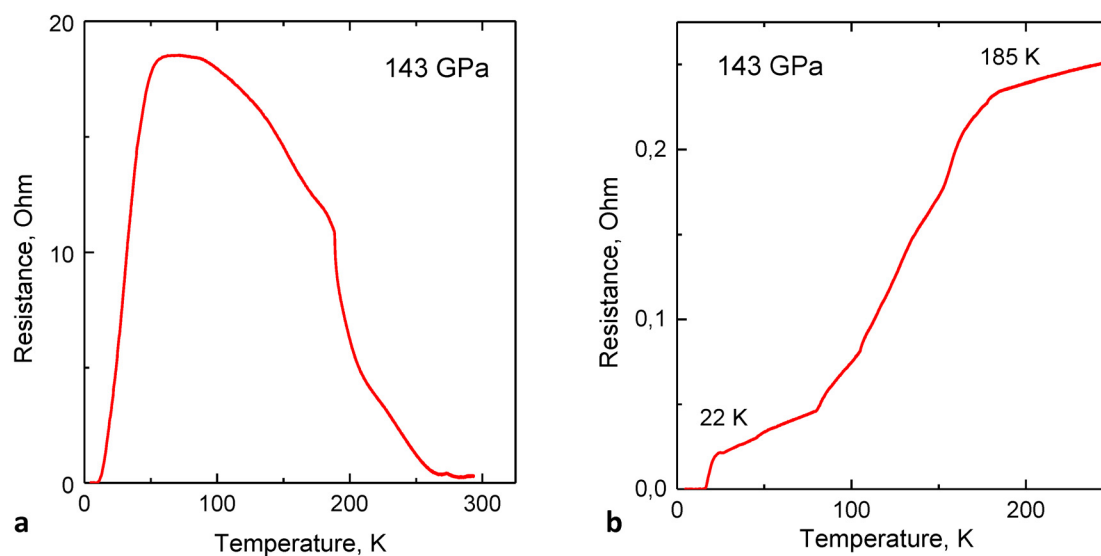
(4) According to Bean's model, the magnetic critical current density of the superconductor can be estimated from the distance between the direct and the returning branches of the magnetic hysteresis loop at a given magnetic field (Fig. 4c). Provided grain radii are about 0.1 μm , the intra-grain critical current J_c is about 10^7 A cm^{-2} .

32. Erements, M. I. Megabar high-pressure cells for Raman measurements. *J. Raman Spectrosc.* **34**, 515–518 (2003).
33. Landau, L. D. & Lifshitz, E. M. *Electrodynamics of Continuous Media* Vol. 8, 1st edn, 173 (Pergamon, 1960).



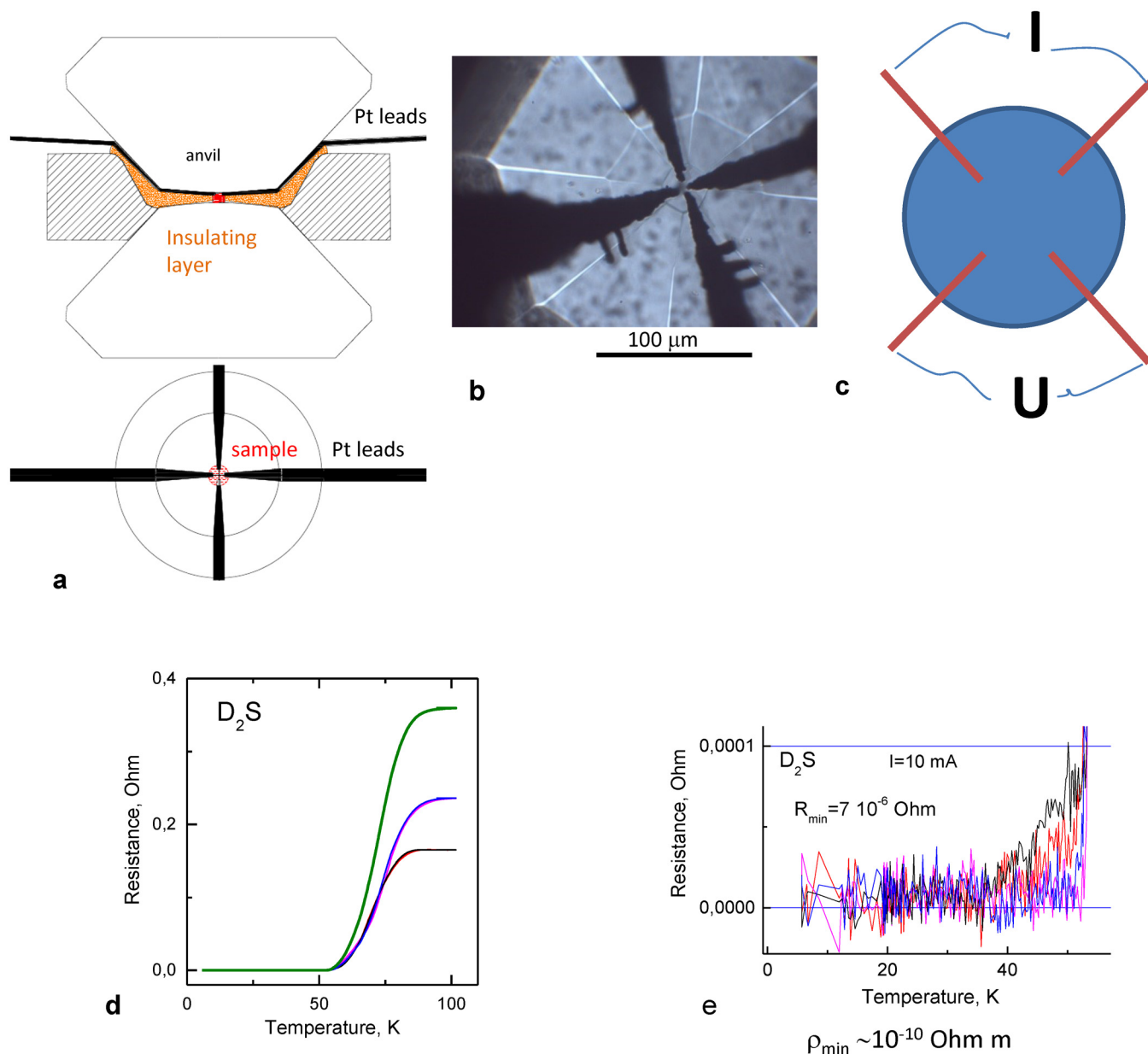
Extended Data Figure 1 | Raman spectra of sulfur hydride at different pressures. **a**, Spectra of sulfur hydride at increasing pressure at ~ 230 K. The spectra are shifted relative to each other. At 51 GPa there is a phase transformation, as follows from disappearance of the characteristic vibron peaks in the $2,100\text{--}2,500\text{ cm}^{-1}$ range. The corresponding spectrum is

highlighted as a bold curve. Bold curves at higher pressure (and the temperature of the measurement) are shown to follow qualitatively the changes of the spectra. The pressure corresponding to the unassigned plots can be determined from the Raman spectra of the stressed diamond anvil³². **b**, Raman spectra of sulfur deuteride measured at $T \approx 170$ K and over the pressure range 1–70 GPa.



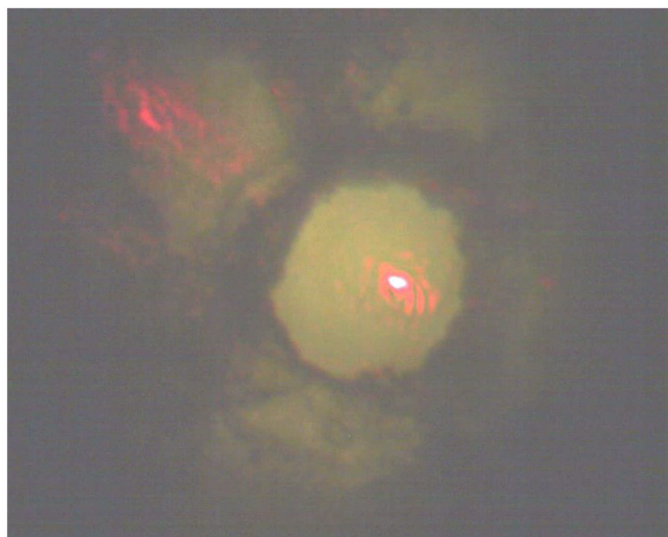
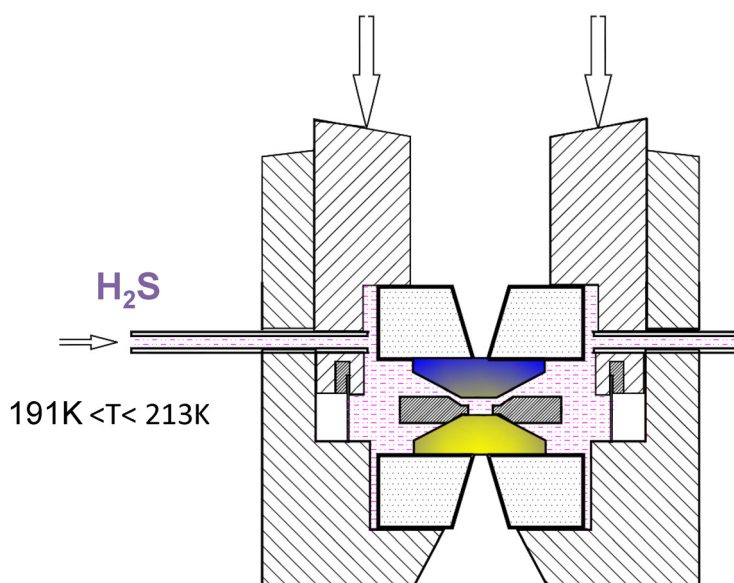
Extended Data Figure 2 | Temperature dependence of the resistance of sulfur hydride at 143 GPa. In this run the sample was clamped in the DAC at $T \approx 200$ K, and the pressure then increased to 103 GPa at this temperature; the further increase of pressure to 143 GPa was at ~ 100 K. **a**, After next cooling to ~ 15 K and subsequent warming, a superconducting transition with $T_c \approx 60$ K was observed, then the resistance strongly decreased with increasing temperature. After successive cooling and warming (**b**; only the warming curve

is shown) a kink at 185 K appeared, indicating the onset of superconductivity. The superconducting transition is very broad: resistance dropped to zero only at ~ 22 K. There are apparent 'oscillations' on the slope. Their origin is not clear, though they probably reflect inhomogeneity of the sample in the transient state before complete annealing. Similar 'oscillations' have also been observed for other samples (see, for example, figure 3 in the Supplementary Information of ref. 9).



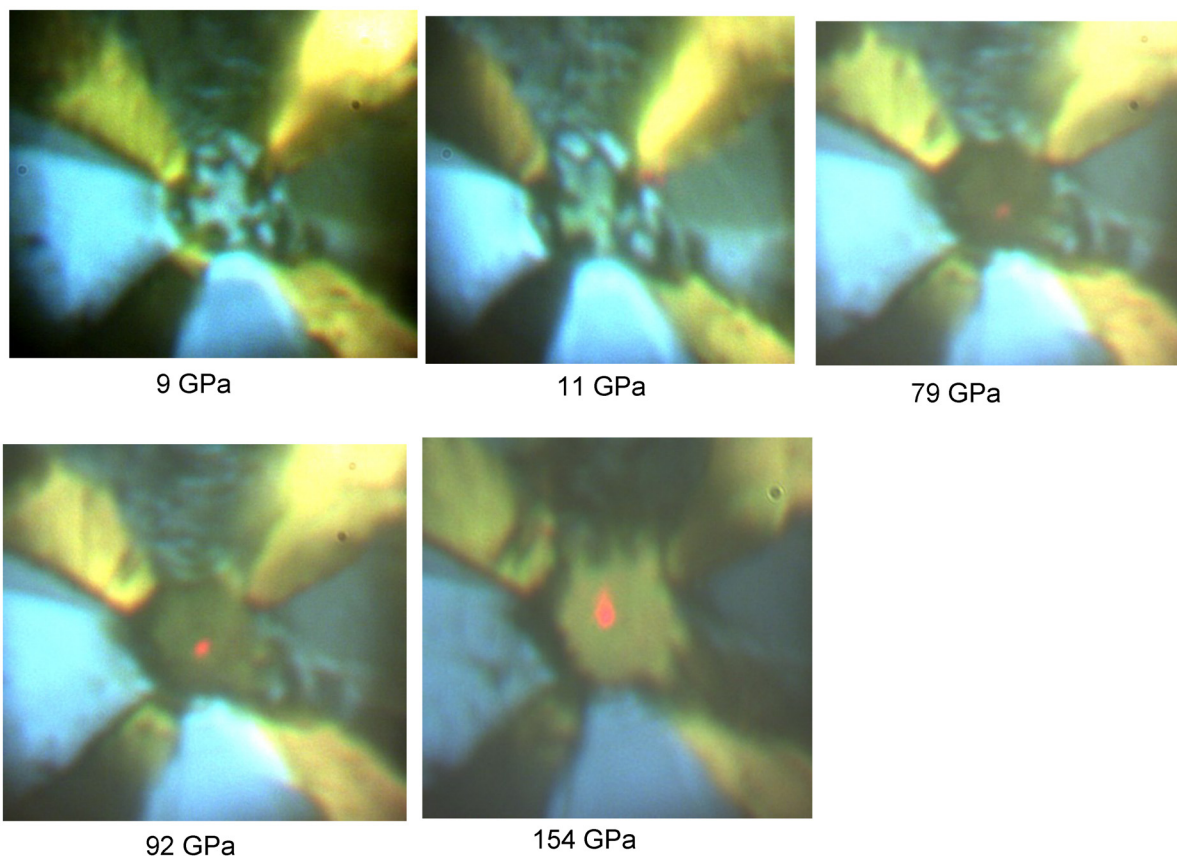
Extended Data Figure 3 | Electrical measurements. **a**, Schematic drawing of diamond anvils with electrical leads separated from the metallic gasket by an insulating layer (shown orange). **b**, Ti electrodes sputtered on a diamond anvil shown in transmitted light. **c**, Scheme of the van der Pauw measurements: current leads are indicated by I , and voltage leads as U . **d**, Typical superconducting step measured in four channels (for different combinations of

current and voltage leads shown in **c**). A sum resistance obtained from the van der Pauw formula is shown by the green line. Note here that the superconducting transition was measured with the un-annealed sample⁹. After warming to room temperature and successive cooling, T_c should increase. **e**, Residual resistance measured below the superconducting transition (**d**). R_{\min} and ρ_{\min} are averaged over four channels shown by different colours.



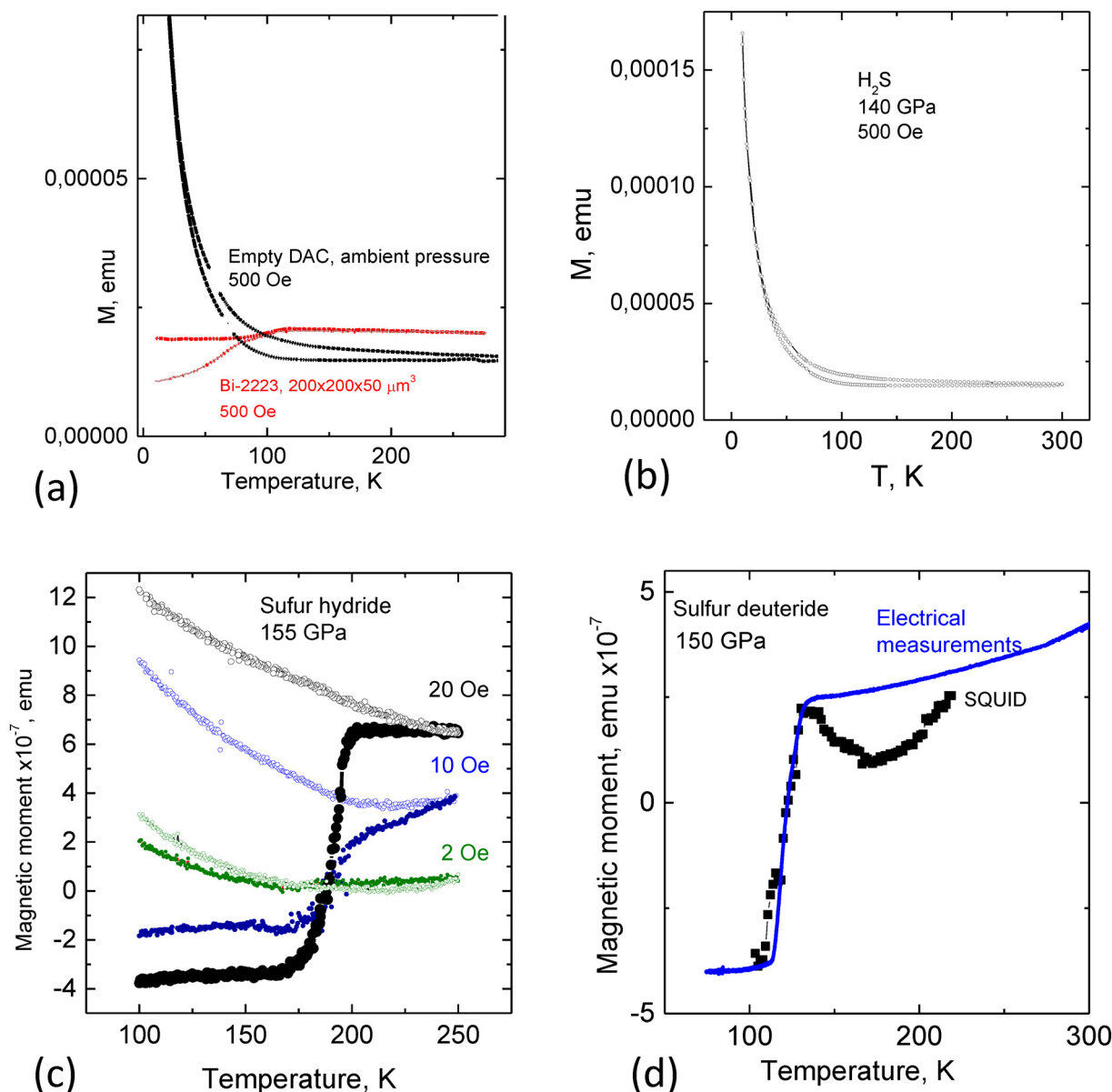
Extended Data Figure 4 | Loading of H₂S. Gaseous H₂S is passed through the capillary into a rim around the diamond anvils (upper panel). When the sample liquefies, in the temperature range 191 K < *T* < 213 K, it is clamped. The process of loading is shown on a video (<https://vimeo.com/131914556>) and a still is shown here (lower panel). On the video, the camera is looking through a hole in the transparent gasket (CaSO₄), and shows a view through the

diamond anvil. At *T* ≈ 200 K, the line to the H₂S gas cylinder was opened and the gas condensed. At this moment, the picture changes due to the different refractive index of H₂S. The second anvil with the sputtered electrodes was then pushed forward, and the hole was clamped. The sample changed colour during the next application of pressure. The red point is from the focused HeNe laser beam.



Extended Data Figure 5 | View of D_2S sample with electrical leads and transparent gasket ($CaSO_4$) at different pressures. The D_2S is in the centre of these photographs, which were taken in a cryostat at 220 K with mixed illumination, both transmitted and reflected. Under this illumination, the

insulating transparent gasket shows blue, and the electrodes yellow. The red spot is the focused HeNe laser beam. The sample, which is initially transparent, becomes opaque and then reflective as pressure is increased.



Extended Data Figure 6 | Magnetic susceptibility measurements with a SQUID. A typical sample (Fig. 4) has a disk shape (diameter 50–100 μm and thickness of few micrometres). In the superconductive state the magnetic moment for this disk is estimated as $M(\text{disk}) \approx 0.2r^3H$ (ref. 33). For a disk of radius $r = 40 \mu\text{m}$ (a sample size typical for DACs in the megabar range) and $H = 2 \text{ mT}$ the expected diamagnetic signal, $M(\text{disk})$ is estimated as $2.6 \times 10^{-7} \text{ emu}$. This value is well above the sensitivity of the SQUID which is $\sim 10^{-8} \text{ emu}$ and, therefore, the signal can be detected. A high-pressure DAC made of Cu:Ti alloy has its own magnetic background signal (a) which increases sharply at low temperatures due to residual paramagnetic impurities. Signal from a large superconducting sample (for example, a Bi-2223 superconductor)

could still be detected without magnetic background subtraction. However, the sulfur hydride sample is not seen (b) unless background has been subtracted (c, d). The background signal acquired in the normal state immediately above T_{onset} has been used for subtraction over all the temperature range taking into account that the magnetic moment of the DAC is fairly temperature independent above 100 K. c, Magnetic measurements for the sample of sulfur hydride at different magnetic fields (labels on curves). The data on sulfur deuteride (d) are compared with the superconducting transition in resistivity measurements (blue curve) which has been scaled to fit the susceptibility data (black points).

Negative refractive index and acoustic superlens from multiple scattering in single negative metamaterials

Nadège Kaina¹, Fabrice Lemoult¹, Mathias Fink¹ & Geoffroy Lerosey¹

Metamaterials, man-made composite media structured on a scale much smaller than a wavelength, offer surprising possibilities for engineering the propagation of waves^{1–6}. One of the most interesting of these is the ability to achieve superlensing—that is, to focus or image beyond the diffraction limit⁷. This originates from the left-handed behaviour—the property of refracting waves negatively—that is typical of negative index metamaterials^{8–10}. Yet reaching this goal requires the design of ‘double negative’ metamaterials, which act simultaneously on the permittivity and permeability in electromagnetics^{11,12}, or on the density and compressibility in acoustics; this generally implies the use of two different kinds of building blocks^{13,14} or specific particles presenting multiple overlapping resonances^{15–17}. Such a requirement limits the applicability of double negative metamaterials, and has, for example, hampered any demonstration of subwavelength focusing using left-handed acoustic metamaterials¹⁸. Here we show that these strict conditions can be largely relaxed by relying on media that consist of only one type of single resonant unit cell. Specifically, we show with a simple yet general semi-analytical model that judiciously breaking the symmetry of a single negative metamaterial is sufficient to turn it into a double negative one. We then demonstrate that this occurs solely because of multiple scattering of waves off the metamaterial resonant elements, a phenomenon often disregarded in these media owing to their subwavelength patterning. We apply our approach to acoustics and verify through numerical simulations that it allows the realization of negative index acoustic metamaterials based on Helmholtz resonators only. Finally, we demonstrate the operation of a negative index acoustic superlens, achieving subwavelength focusing and imaging with spot width and resolution 7 and 3.5 times better than the diffraction limit, respectively. Our findings have profound implications for the physics of metamaterials, highlighting the role of their subwavelength crystalline structure, and hence entering the realm of metamaterial crystals. This widens the scope of possibilities for designing composite media with novel properties in a much simpler way than has been possible so far.

Negative index electromagnetic materials, as predicted in ref. 11, can be designed by achieving simultaneously negative permittivity ϵ and negative permeability μ , although such materials remained theoretical until metamaterials were proposed^{1,2}. Metamaterials are subwavelength scaled composite media in which the excitation field as well as the response of the medium’s unit cells can be averaged, so that they can be described in terms of effective parameters (ϵ_{eff} or μ_{eff}), which can be negative. It then becomes straightforward to achieve negative index metamaterials by combining two building blocks, each bringing either negative ϵ or negative μ in electromagnetism or equivalently negative density ρ or negative compressibility χ in acoustics, hence potentially allowing superlensing⁷. Another approach to the design of negative index materials is possible, in which two overlapping resonances in the same frequency range are used, one being a monopolar resonance and the other one being dipolar. These resonances can originate from the same element¹⁵, a Mie scatterer for instance. But they can also arise from elements composed of two

building blocks strongly coupled via inductive or capacitive effects, and whose symmetry has been broken^{19,20}. This approach is, however, not universal since these strong near-field couplings apply only to specific geometries and materials. On the other hand, photonic/phononic crystals can also be left-handed, this time only because of multiple scattering in a periodic structure that leads to a band folding, thus offering negative refraction^{21,22}. Those crystals based on Bragg interferences present, however, the major drawback of having a spatial period that is comparable to the wavelength, which makes subwavelength resolution difficult to achieve. We now show that, as in crystals, multiple scattering has profound consequences in resonant metamaterials despite their deeply subwavelength scale. This can drastically simplify the design of negative index metamaterials.

For the sake of simplicity, we start our study with the simplest possible metamaterial, that is, a one-dimensional chain of resonators organized periodically on a deeply subwavelength scale a (Fig. 1a). The building block of the unit cell is modelled by a resonant point scatterer, with a resonant frequency f_0 and a linewidth Γ (see Methods). This approach is hence valid for any type of wave (acoustic or electromagnetic) and any type of resonator. The dispersion relation of such a chain, with $a = \lambda_0/12$, is calculated analytically using a combination of a Green’s function formalism and a transfer matrix approach²³ (see Methods). This analytical approach, which includes multiple scattering, is very general and applies so long as strong coupling between resonant elements can be neglected, which is valid for many resonators whatever the spacing, and for any resonator if a minimum separation distance is maintained²⁴. The obtained dispersion is well-known to be polaritonic (Fig. 1a), or equivalently, this medium can be described by a set of two effective parameters (ϵ_{eff} , μ_{eff} in electromagnetism or ρ_{eff} , χ_{eff} in acoustics) with only one of them being negative: this is a so-called single negative metamaterial.

Starting from this unit cell that creates the single negative property, we build two new configurations just by breaking the symmetry in two different ways: either (1) we build a bi-periodic chain by off-centring one resonator out of two or (2) we create a bi-disperse chain by slightly shifting the resonance frequency of every other resonator. For both configurations, we analytically calculate the dispersion relation and extract the corresponding effective refractive index (Fig. 1b, c). Both of those new metamaterials now exhibit, in the bandgap of the single negative medium, a new band of propagating waves that presents a negative index. To confirm the existence of this negative band, we evaluated, for each configuration, the effective parameters retrieved from the transmission and reflection coefficients of a single unit cell and found a negative index of refraction for both (see Methods and Extended Data Fig. 1).

To understand the origin of this negative band, we perform a parametric study of these symmetry-broken formal metamaterials. For the bi-periodic chain, we vary the shift in position, while for the bi-disperse one we vary the frequency detuning between resonators. For each parameter, we analytically evaluate the dispersion relation with our approach that takes multiple scattering into account and extract the effective index of the metamaterials from it. We display

¹Institut Langevin, ESPCI ParisTech and CNRS UMR 7587, 1 rue Jussieu, 75005 Paris, France.

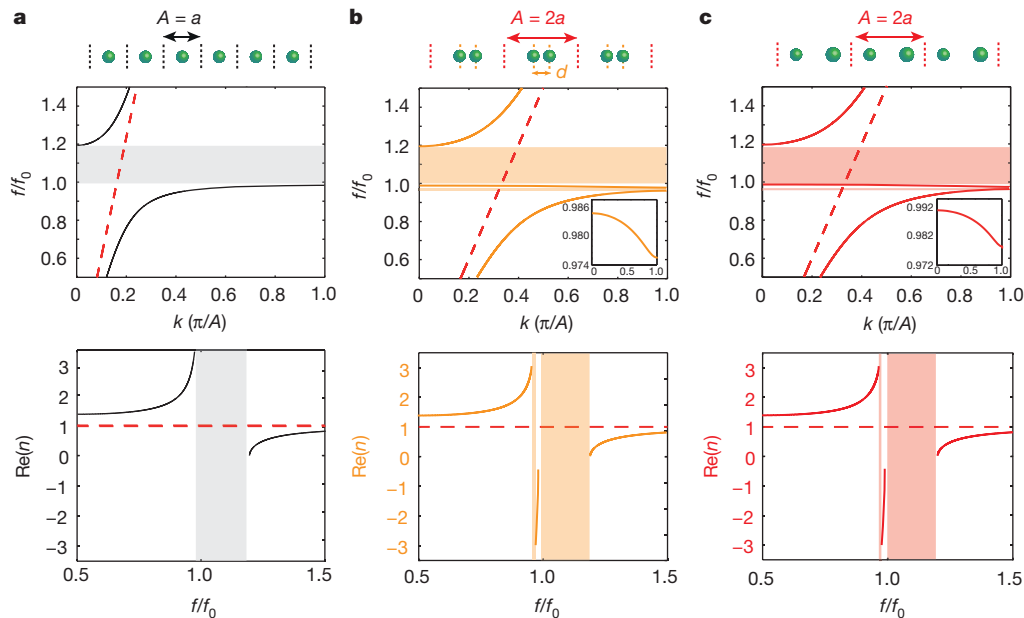


Figure 1 | Analytical study of one-dimensional chains of resonant point scatterers. **a**, Top, schematic view of the regular one-dimensional chain of resonant scatterers (green spheres) resonating at f_0 , with f_0 arbitrarily set to 0.15 GHz. The chain consists of building blocks of period $\lambda_0/12$ (dotted black lines) each containing one resonant scatterer. Middle panel, the corresponding dispersion relation, polaritonic as expected. Bottom panel, the real part of the index. Also shown are free space properties (dashed red lines). Frequencies are normalized by the resonance frequency f_0 of the scatterers while the wavevectors are limited to the first Brillouin zone. The shaded areas indicate

the bandgaps, where the wavevector and index turn imaginary. The same study is conducted for the bi-periodic chain (**b**) with two identical scatterers separated by a distance d (dashed orange lines) per building block (dashed red lines) of period $A = \lambda_0/6$. The distance d is arbitrarily set to $0.8193A$. Inset, an enlargement of the negative band. In the case of the bi-disperse chain (**c**) with two scatterers having different resonance frequencies f_0 and f_1 (represented as spheres with slightly different radii) per unit cell of period $A = \lambda_0/6$, the same behaviour is found. The curves are plotted for $f_1 = 1.0122f_0$. Inset, enlargement of the negative band.

those results in a colour-coded representation (Fig. 2). The bi-periodic chain presents a negative index band, independently of the spacing, which is symmetric with respect to the shift in position. In the bi-disperse case, however, the existence of this negative band depends strongly on the frequency detuning, suggesting that multiple scattering may be involved. Relying on this intuition, we also extract, for the same set of parameters, the effective index in the same way as is frequently

done in the field of metamaterials, namely with an independent scattering approximation (ISA). This time we never retrieve any negative index, but rather two positive index bands typical of a double polariton. This clearly shows that the negative index arises from multiple scattering between the resonators of the unit cell, even if the distance is far below the wavelength. This, in turn, explains why, for the bi-disperse chain, the existence of the negative index depends

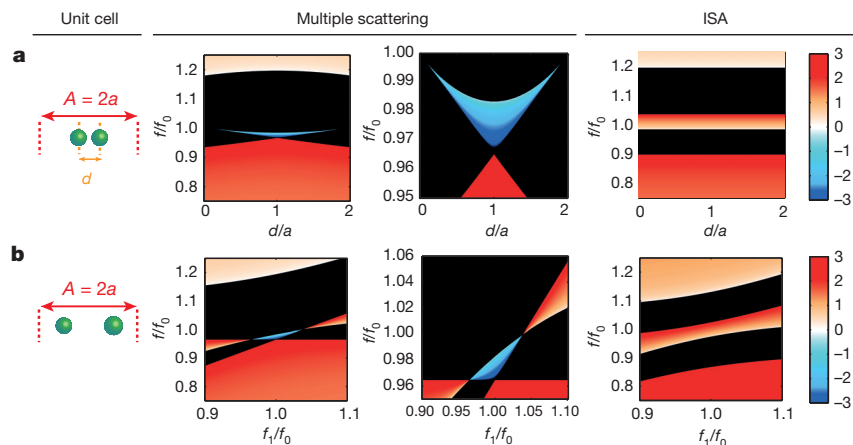


Figure 2 | Importance of multiple scattering within the building block. **a**, Building block of the bi-periodic one-dimensional chain with the two characteristic dimensions d and $A = 2a$ ('unit cell'). Panels under 'multiple scattering' show the index of the bi-periodic chain as a function of the normalized frequency when scanning the distance d from 0 to A , confirming the arising of a negative band (blue) that is logically symmetric across the distance a , corresponding to the regular chain. The black areas indicate the bandgaps while the reddish ones represent the first and third positive index bands of the dispersion relation. Left panel, wide scan; right panel, enlargement of the frequency range where the negative band occurs. Panel under 'ISA'

shows the index for the same chain calculated using the independent scattering approximation (ISA) leading to only positive index bands, uniformly, whatever the distance d . **b**, As for **a** but for the bi-disperse chain while varying the detuning frequency f_1 of the second scatterer, keeping f_0 fixed. Contrarily to the bi-periodic chain, however, the second band is of negative index solely within a certain range of detuning close to f_0 and turns positive for a too large detuning f_1/f_0 . Once again, the independent scattering approximation (right) leads to only positive index bands. Note that the cases of $d = a$ and $f_0 = f_1$ are mathematical band folding artefacts. The colour scales represent the real part of the index of refraction (dimensionless).

strongly on the chosen detuning: for too large a resonance frequency mismatch, the two resonators cannot couple any longer owing to multiple scattering.

To grasp the physics of the approach, we carefully studied the fields created by a dimer (the new unit cell of the symmetry-broken media) and made the following observation: multiple scattering creates a dipolar resonance (the two resonators are out-of-phase) overlapping with a monopolar resonance of the dimer. This dipolar resonance is responsible for the opening of a narrow transparency window within the large out-of-phase response of the monopolar resonance of the two resonators constituting the dimer. This is analogous to electromagnetic induced transparency in quantum physics²⁵, or more precisely to its metamaterial equivalents²⁶. We stress that this dipolar resonance results from multiple scattering, and that the conventional homogenization procedure based on the independent scattering approximation cannot retrieve it. Moving from the unit cell to an infinite medium, this dipolar mode gives rise to a band of propagating waves within the bandgap of the single negative medium, the latter being a consequence of the monopolar resonance of the unit cell. This band has a negative slope, or equivalently, the metamaterial now presents a negative index. This originates physically from the fact that owing to the symmetry breaking, the lower polaritonic band folds in the first Brillouin zone, analogous to optical branches in diatomic crystals or band folding in

photonic or phononic crystals. Here, though, the band folding, owing to the change of sign of the Bloch mode between the two edges of the unit cell, has a different origin in those metamaterials compared to Bragg interference based crystals. Indeed, while in the latter it arises from the fact that the host medium wavelength becomes smaller than twice the lattice constant, in the former the change of sign results from the dipolar nature of the resonant mode within the unit cell. This implies that, contrary to negative refraction in crystals, this new phenomenon exists even if the scale of the metamaterial is deeply subwavelength, and happens at the same frequency as the resonance of the original building block. As a consequence, while negative refraction in photonic/phononic crystals is inherently diffraction limited, our concept permits subdiffraction resolution, as we will show later on. Furthermore, since the original single negative effective property does not rely on spatial order²⁷, this negative index band should be robust even in a metamaterial constituted of randomly placed dimers. Indeed, both the monopolar and the multiple-scattering-induced dipolar resonances should remain, hence leading to a negative index medium¹⁵. We finally note that the two symmetry-broken metamaterials are strictly equivalent under a transformation approach^{28,29}.

Since our idea is very general, we apply it to acoustics, designing a two-dimensional acoustic metamaterial whose unit cell is made of soda (soft drink) cans, that is, acoustic Helmholtz resonators^{24,30}. First, to

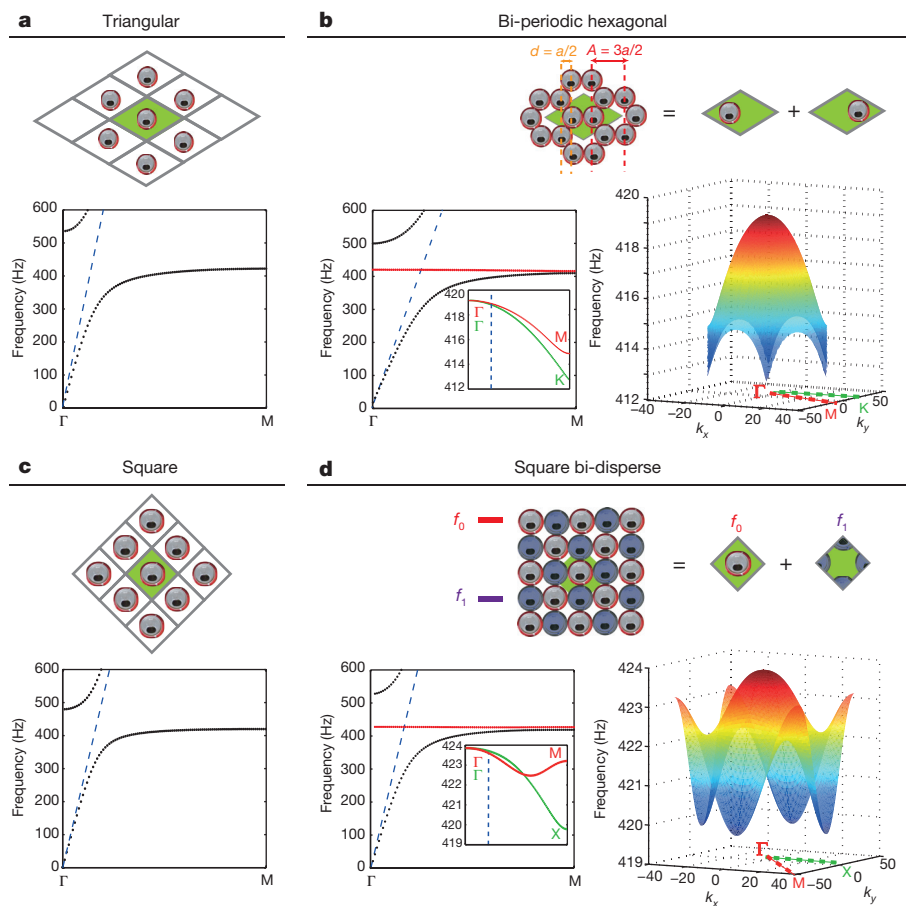


Figure 3 | Numerical simulations of soda-can two-dimensional regular and symmetry-broken media. **a**, Main panel, dispersion relation in the ΓM direction (dotted black lines) of a regular triangular array along with free space dispersion (dashed blue line); top panel, schematic of the medium and building block with one can per unit cell (shaded green). **b**, As **a** but for a corresponding symmetry-broken medium, that is, the bi-periodic honeycomb array but with the building block now consisting of two cans. This medium can be decomposed on two triangular lattices with cans slightly off-centred. The negative index band in the dispersion relation is plotted in red (left panel). Inset,

enlargement of the negative band in the ΓM (red) and ΓK (green) directions. In addition, a surface plot of the negative band over the whole Brillouin zone is shown (right panel). **c**, **d**, As for **a**, **b** but for the regular square lattice and the corresponding symmetry-broken medium, that is, the bi-disperse square lattice. The building block of the bi-disperse square lattice can be shown schematically as a superposition of two regular square lattices: the first with one can (f_0) in the centre, and the second with one quarter of a can displayed in each corner (f_1). The specific directions of the square Brillouin zone are ΓM (red) and ΓX (green).

realize the bi-periodic two-dimensional metamaterial, we use the so-called honeycomb lattice, which exhibits this double periodicity in any of the ΓM directions. This ‘crystal’ is made of a diamond-like unit cell consisting of two resonators, and is compared to the triangular lattice which has the same unit cell but with only one resonator. Numerical simulations using Comsol Multiphysics gives the dispersion of both the regular and the symmetry-broken lattices (Fig. 3a, b). The triangular lattice medium presents a polaritonic dispersion relation, or equivalently it can be modelled as a single negative medium, while the honeycomb lattice (which actually consists of the superposition of two identical triangular lattice crystals) displays a negative band. The dispersion depends slightly on the propagation direction but remains rather isotropic as shown in the surface plot (Fig. 3b) and can thus be described with an isotropic negative effective index of refraction. For the bi-disperse two-dimensional lattice, we mix two square lattices of slightly detuned Helmholtz resonators in order to build a new square lattice whose unit cell contains two resonators (Fig. 3c, d). This bi-disperse resonant crystal exhibits a negative branch although it is simply the superposition of two almost identical single negative media. In this case, the propagation is less isotropic, since the geometry of a square unit cell tends to deform the isofrequency contours near the corners of the first Brillouin zone. There is, however, no doubt that one can find a more isotropic medium. An effective medium approach applied to the specific case of our simulated soda-can unit cells (Extended Data Fig. 2) leads to the same results.

To show the potential of our approach, we experimentally demonstrate a negative index acoustic metamaterial superlens. We work with the bi-periodic medium, here a honeycomb arrangement of soda

cans, each one resonating at $f_0 = 420$ Hz ($\lambda_0 \approx 80$ cm). We build a slab with 124 cans (Fig. 4a), surrounded by acoustic absorbers to avoid reflections of sound off the boundaries of the room. An 8-cm-wide loudspeaker located approximately 5 cm away from the input interface of the medium is used as the source of sound while microphones mounted on a two-dimensional translational stage measure the acoustic field above the lens. We work at $f = 417.5$ Hz, a frequency at the lower edge of the negative band, that is, where the effective negative index norm is the highest, in order to get the best possible resolution. We observe that the field map displays a cone characteristic of negative refraction in dissipative media (Fig. 4b). By compensating for the losses that occur during the propagation in the lens (losses that are experimentally characterized for the whole frequency range of the negative band in Extended Data Fig. 3), we can clearly distinguish the path for sound refraction, with a focal spot inside the lens (Fig. 4c), in very good agreement with the Snell–Descartes law for a metamaterial with an effective index around -3 , consistent with our analytical and numerical results (see Methods). On the other side of the superlens, in the vicinity of the surface, we record the image of the source with a $\lambda_0/15$ full-width at half-maximum. This is much smaller than the diffraction limited focus obtained without the lens (black curve in Fig. 4d), and even smaller than the width of the source, $\lambda_0/5$, owing to a hotspot created by the aperture of a single soda can (a small hole of $\lambda/15$ width). We stress that these focusing results are without doubt obtained owing to negative refraction rather than being due to a canalization effect that would arise from a flat band, for instance. This is further confirmed by simulations of negatively refracted Gaussian beams impinging on a larger soda-can slab, by

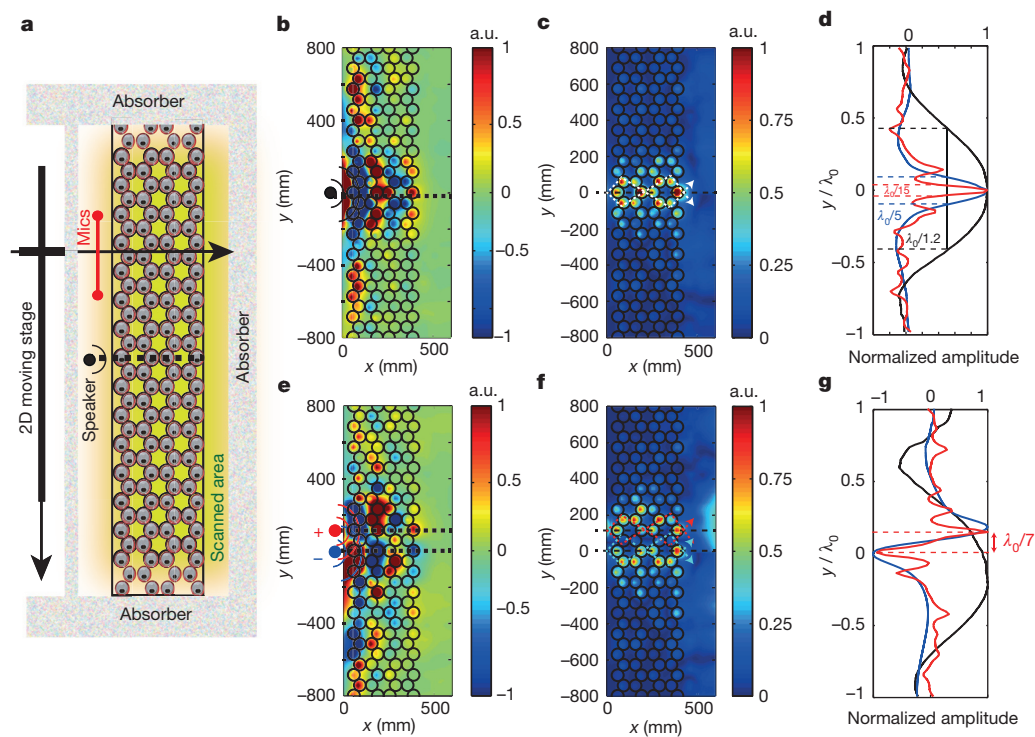


Figure 4 | Experimental demonstration of subwavelength focusing and imaging using a flat acoustic superlens. **a**, The flat lens, composed of a compact hexagonal array of soda cans, is supplied with sound by a loudspeaker (speaker) centred close to the surface of the medium. Two microphones (mics) mounted on a stage moving in two dimensions record the acoustic pressure field less than 1 cm away from the tops of the cans over the whole scanned area. Absorbers surround the lens to prevent undesired reflections. **b**, The real part of the pressure field (colour coded in a.u., arbitrary units) at $f = 417.5$ Hz, which is near the edge of the negative band, is displayed on a map. **c**, Map of the pressure intensity field after compensation of losses occurring during

the propagation within the lens; the directions of the refracted waves are plotted in white dashed lines, clearly displaying the features of negative refraction. **d**, The normalized amplitude of the field in the close vicinity of the output surface proves the existence of a focusing area of $\lambda_0/15$ (red line) while the source (blue line) is $\lambda_0/5$ wide and the control experiment (black line), that is, without the lens, is $\lambda_0/1.2$ wide. **e–g**, As **b–d** but for two sources that play sound out of phase to demonstrate super-resolution. It clearly demonstrates the same negative refraction results with a resolution of $\lambda_0/7$. The dashed red (blue) line in **f** represents the direction of the refracted waves coming from the first (second) source.

effective medium simulations and by experimental observations of frequency dependent foci positions (see Methods and Extended Data Figs 4–8).

Super-resolution can also be demonstrated by being able to discriminate two sources separated by less than half a wavelength. Two loudspeakers, emitting out-of-phase, were placed near the input surface, separated by 13 cm ($\lambda_0/7$). The measured pressure field, as well as the loss-compensated field maps (Fig. 4e, f), show that the slab produces two distinguishable foci inside the superlens. In the focal plane, the two images are efficiently separated, thereby demonstrating a $\lambda_0/7$ imaging resolution, far beyond the diffraction limit, contrary to the control experiment realized without the superlens (black curve in Fig. 4g). We have further verified that the two sources can be distinguished whatever the phase shift between them (Extended Data Fig. 9).

We have demonstrated that it is fairly easy to build double negative media from single negative ones. By breaking the symmetry of the unit cell of a single negative medium (either by changing the spacing or by frequency detuning), multiple scattering of waves guarantees the existence of an overlap between a dipolar resonance and a monopolar one. This eventually results in a negative effective index of refraction when considering an infinite medium. This approach is very general as long as near-field coupling between resonators can be neglected, and brings a new paradigm to the physics of metamaterials since multiple scattering is often neglected owing to the subwavelength spatial scale of those media. We emphasize that such a negative effective index of refraction should be insensitive to the random positioning of dimers, and that it should be easily transposable to three-dimensional metamaterials. Our results open the way towards metamaterial crystals.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 20 February; accepted 12 June 2015.

- Pendry, J., Holden, A., Stewart, W. & Youngs, I. Extremely low frequency plasmons in metallic mesostructures. *Phys. Rev. Lett.* **76**, 4773–4776 (1996).
- Pendry, J. B., Holden, J., Robbins, D. J. & Stewart, W. J. Magnetism from conductors and enhanced nonlinear phenomena. *IEEE Trans. Microw. Theory Tech.* **47**, 2075–2084 (1999).
- Liu, Z. *et al.* Locally resonant sonic materials. *Science* **289**, 1734–1736 (2000).
- Deymier, P. A. *Acoustic Metamaterials and Phononic Crystals* (Springer, 2013).
- Engheta, N. & Ziolkowski, R. *Metamaterials: Physics and Engineering Explorations* (Wiley & Sons, 2006).
- Craster, R. V. & Guenneau, S. *Acoustic Metamaterials: Negative Refraction, Imaging, Lensing and Cloaking* (Springer, 2012).
- Pendry, J. B. Negative refraction makes a perfect lens. *Phys. Rev. Lett.* **85**, 3966–3969 (2000).
- Lee, S. H., Park, C. M., Seo, Y. M., Wang, Z. G. & Kim, C. K. Composite acoustic medium with simultaneously negative density and modulus. *Phys. Rev. Lett.* **104**, 054301 (2010).
- Liang, Z., Feng, T., Lok, S., Liu, F. & Ng, K. Space-coiling metamaterials with double negativity and conical dispersion. *Sci. Rep.* **3**, 1–6 (2013).
- Xie, Y., Popa, B. I., Zigoneanu, L. & Cummer, S. a. Measurement of a broadband negative index with space-coiling acoustic metamaterials. *Phys. Rev. Lett.* **110**, 175501 (2013).
- Veselago, V. G. The electrodynamics of substances with simultaneous negative values of ϵ and μ . *Phys. Uspekhi* **10**, 509–514 (1968).
- Smith, D. R., Padilla, W. J., Vier, D. C., Nemat-Nasser, S. C. & Schultz, S. Composite medium with simultaneously negative permeability and permittivity. *Phys. Rev. Lett.* **84**, 4184–4187 (2000).
- Shelby, R. A., Smith, D. R. & Schultz, S. Experimental verification of a negative index of refraction. *Science* **292**, 77–79 (2001).
- Guenneau, S., Movchan, A., Pétursson, G. & Anantha Ramakrishna, S. Acoustic metamaterials for sound focusing and confinement. *New J. Phys.* **9**, 399 (2007).
- Li, J. & Chan, C. T. Double-negative acoustic metamaterial. *Phys. Rev. E* **70**, 055602 (2004).
- Brunet, T. *et al.* Soft 3D acoustic metamaterial with negative index. *Nature Mater.* **14**, 384–388 (2015).
- Zhu, R., Liu, X. N., Hu, G. K., Sun, C. T. & Huang, G. L. Negative refraction of elastic waves at the deep-subwavelength scale in a single-phase metamaterial. *Nature Commun.* **5**, 5510 (2014).
- Zhang, S., Yin, L. & Fang, N. Focusing ultrasound with an acoustic metamaterial network. *Phys. Rev. Lett.* **102**, 194301 (2009).
- Kanté, B. *et al.* Symmetry breaking and optical negative index of closed nanorings. *Nature Commun.* **3**, 1180 (2012).
- Christ, A., Martin, O. J. F., Ekinci, Y., Gippius, N. A. & Tikhodeev, S. G. Symmetry breaking in a plasmonic metamaterial at optical wavelength. *Nano Lett.* **8**, 2171–2175 (2008).
- Sukhovich, A. *et al.* Experimental and theoretical evidence for subwavelength imaging in phononic crystals. *Phys. Rev. Lett.* **102**, 154301 (2009).
- Notomi, M. Theory of light propagation in strongly modulated photonic crystals: refractionlike behavior in the vicinity of the photonic band gap. *Phys. Rev. B* **62**, 10696–10705 (2000).
- Yariv, A. & Yeh, P. *Optical Waves in Crystals* (Wiley, 1984).
- Lemoult, F., Kaina, N., Fink, M. & Lerosey, G. Wave propagation control at the deep subwavelength scale in metamaterials. *Nature Phys.* **9**, 55–60 (2013).
- Fleischhauer, M. & Marangos, J. P. Electromagnetically induced transparency: optics in coherent media. *Rev. Mod. Phys.* **77**, 633–673 (2005).
- Papasimakis, N., Fedotov, V., Zheludev, N. & Prosvirnin, S. Metamaterial analog of electromagnetically induced transparency. *Phys. Rev. Lett.* **101**, 253903 (2008).
- Kaina, N., Lemoult, F., Fink, M. & Lerosey, G. Ultra small mode volume defect cavities in spatially ordered and disordered metamaterials. *Appl. Phys. Lett.* **102**, 144104 (2013).
- Leonhardt, U. Optical conformal mapping. *Science* **312**, 1777–1780 (2006).
- Pendry, J. B., Schurig, D. & Smith, D. R. Controlling electromagnetic fields. *Science* **312**, 1780–1782 (2006).
- Lemoult, F., Fink, M. & Lerosey, G. Acoustic resonators for far-field control of sound on a subwavelength scale. *Phys. Rev. Lett.* **107**, 064301 (2011).

Acknowledgements We thank A. Lagendijk for sharing his knowledge of multiple scattering techniques. This work was supported by LABEX WIFI (Laboratory of Excellence within the French Program “Investments for the Future”) grants ANR-10-LABX-24 and ANR-10-IDEX-0001-02 PSL*, and by Agence Nationale de la Recherche grant ANR-13-JS09-0001-01. N.K. acknowledges funding from the French “Direction Générale de l’Armement”.

Author Contributions G.L. conceived the original idea. G.L. and F.L. supervised the project. N.K. and F.L. performed the analytical and numerical work. N.K., F.L. and G.L. performed the experiments, analysed the data, and wrote the manuscript. All authors discussed the results and commented on the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to G.L. (geoffroy.lerosey@espci.fr).

METHODS

Analytical calculations. To calculate the dispersion relation of an infinite periodic and lossless medium of lattice constant a , one only needs to know the transmission coefficient $T(\omega)$ through a single unit cell²⁴. Indeed, the Bloch wavenumber k is the solution of the equation:

$$\cos(ka) = \text{Re} \left(\frac{e^{-ik_0 a}}{T(\omega)} \right)$$

Evaluating $T(\omega)$ is usually performed by the use of simulations, but here we propose to use an analytic formulation by using resonant point scatterers³¹. In this context, a resonator is described by a so-called t -matrix $t(\omega)$. The frequency dependence of $t(\omega)$ has to satisfy the optical theorem (which is equivalent to energy conservation), and for a resonant point scatterer in one dimension it takes the form:

$$t(\omega) = \frac{\omega}{c} \times \frac{\Gamma}{\omega_0 - \omega - i\frac{\Gamma}{2}}$$

where ω_0 is the resonating pulsation of the point scatterer, and Γ is the resonance linewidth. If we consider that this point scatterer, situated at the coordinate point $x = x_0$, is excited by a normalized incoming plane wave, we can calculate the field at any position with the formula:

$$\psi(x, \omega) = e^{ik_0 x} + G_0(x - x_0) t(\omega) e^{ik_0 x_0}$$

where $G_0(x)$ is the one-dimensional Green's function. As a consequence, the transmission coefficient through a single point scatterer in a one-dimensional space is:

$$T(\omega) = 1 + \frac{ic}{2\omega} t(\omega)$$

To calculate the transmission coefficient for the case of unit cells containing N different resonators, we need to invoke multiple scattering theory³¹. Considering N resonators (the resonator labelled α is located at coordinate x_α and is described with the t -matrix $t_\alpha(\omega)$), we can calculate the field at any position in response to a normalized plane wave excitation with:

$$\psi(x, \omega) = e^{ik_0 x} + \sum_{\alpha=1}^N \sum_{\beta=1}^N G_0(x - x_\alpha) W_{\alpha\beta}(\omega) e^{ik_0 x_\beta}$$

Here we have introduced the matrix $W(\omega)$, the inverse elements of which are:

$$(W_{\alpha\beta}(\omega))^{-1} = \delta_{\alpha\beta} (t_\alpha(\omega))^{-1} - (1 - \delta_{\alpha\beta}) G_0(x_\alpha - x_\beta)$$

Solving the multiple scattering problem for a two-resonator unit cell in order to get the transmission coefficient $T(\omega)$ is thus only a problem of inverting the 2×2 matrix and this is what we performed frequency by frequency in order to build the dispersion relations shown in Fig. 2.

The multiple scattering theory is well-known in the field of random media, where numerous scatterers are considered. Inverting the matrix can become difficult and it is sometimes useful to neglect some terms in the multiply scattered waves in order to describe the coherent wave that travels inside the medium. One approach, namely the independent scattering approximation, consists of neglecting the multiply scattered terms that see a given scatterer at least twice. For the two-resonator unit cell, the independent scattering approximation reduces the summation, and the transmission coefficient in this approximation is:

$$T_{\text{ISA}}(\omega) = 1 + \frac{ic}{2\omega} t_1(\omega) + \frac{ic}{2\omega} t_2(\omega) - \frac{c^2}{2\omega^2} t_1(\omega) t_2(\omega) e^{ik_0(x_2 - x_1)}$$

Choice of parameters for the analytical, numerical and experimental results.

For the semi-analytical study, the distance between the point scatterers was set to $\lambda_0/12$, where λ_0 is the wavelength at resonance. This distance corresponds to the simulated and experimental typical size of the unit cell on soda-can honeycomb lattices, when cans are packed in a compact arrangement. The linewidth of the scatterers' resonance (equivalently the quality factor of the resonator) was taken to be of the same order as in the experiment. This allows us to model the refracted beams with an index of $n = -3$ (corresponding to wavevectors at the edge of the Brillouin zone of a $2a = \lambda/6$ unit cell); this allows a good fit to the experimental data at the bottom band edge frequency $f = 417.5$ Hz.

Effective parameters retrieval for media made of resonant point scatterers. In order to describe the propagation in media presenting a negative index of refraction, a common approach is to use the effective parameters. In the main text and in Fig. 1, we only presented the dispersion relations of the studied media and extracted an effective index of refraction. Here, to show the generality of our

approach, we retrieve the effective parameters with commonly used procedures³². At this stage of this Letter, our demonstration is based on the simplest object that can be used to build the resonant metamaterial, that is, a resonant point scatterer, so that it remains valid for both electromagnetic and acoustic waves. Since the experimental proof of concept is, however, realized with the specific case of acoustic waves, we here extract parameters corresponding to acoustics, namely the effective compressibility χ , the effective mass density ρ and the corresponding effective phase velocity c , all normalized by the parameters of the host medium, which in this case is air. The exact same solutions can be, however, found for electromagnetic waves with effective relative permittivity ϵ , effective relative permeability μ and effective index of refraction.

To evaluate those effective parameters, we first calculate the resonant transmission coefficient $T(\omega)$ and the resonant reflection coefficient $R(\omega)$ in the studied one-dimensional chain configuration for the three media presented in Fig. 2; a periodic medium made of a single resonator, the so-called bi-periodic medium and finally the bi-disperse medium consisting of slightly detuned resonators. In these three cases the non-radiative losses of each point scatterer are neglected. From those coefficients, we use the method presented in ref. 32 that has the advantage of avoiding the uncertainty in the sign of the solutions. The retrieved effective coefficients are presented in Extended Data Fig. 1. This unambiguously demonstrates that the negative bands in Fig. 1 are due to the double negativity of the two effective parameters describing the medium. Indeed, the two shaded areas in Extended Data Fig. 1a and b correspond to the part of the spectrum where the effective phase velocity, the effective compressibility and the effective mass density are all simultaneously negative. Nevertheless, such effective parameters, even if commonly used in the community, present some weird behaviours that are due to periodic effects. In the case of the single negative medium for instance, the effective density displays a dip which is often referred to as an anti-resonance and which we do not think is really relevant.

3D simulations. *Description.* All the simulations are performed using the eigen-solver of the finite element simulation software Comsol Multiphysics. The unit cell of each medium (highlighted in green in Fig. 3) is simulated by applying Bloch periodic boundary conditions to retrieve the dispersion relation. The soda cans are modelled with rigid-walled cylinders of 66-mm diameter. The simulated soda can has a resonance frequency around 418 Hz, slightly below the experimental resonance of 420 Hz due to a small difference in the bottom of their geometry. To create a resonator with a small detuning in the case of the bi-disperse medium, we reduce the height of the simulated can by 3.5 mm, which is equivalent to filling a can with 1 cl of water³⁰. The distance a is set to the diameter of the cans (66 mm, that is, approximately $\lambda_0/12$). The simulated triangular and honeycomb crystals have a unit cell with a diamond shape with diagonals a and $\sqrt{3}a$. The simulated square unit cells in the case of the bi-disperse medium have a diagonal of $A = 2a$. *Effective parameters retrieval.* We then apply the same procedure to retrieve the effective parameters of the acoustic metamaterials made of soda cans presented in the main text (Fig. 3). We perform three-dimensional simulations using Comsol Multiphysics to get the transmission coefficients through a single unit cell of the metamaterials in the three different configurations. The reference metamaterial, which behaves as a single negative medium, consists of a triangular arrangement of identical soda cans with a lattice constant of $3a/2$ (Fig. 3a). We evaluate the plane wave transmission through one layer of soda cans and apply the method of ref. 32 to retrieve the effective parameters describing the medium. The results displayed in Extended Data Fig. 2 shows that such a medium behaves as a single negative medium, since only the effective compressibility falls below zero. When some losses comparable to the experimental ones are added to the simulation, we see that this metamaterial keeps its single negative property. We then simulated the transmission through the honeycomb arrangement of cans (the two-dimensional equivalent of the bi-periodic chain). In this case, the simulation consists of measuring the transmission and reflection on two cans separated by a distance of $a/2$ along the propagation direction and $\sqrt{3}a/2$ in the transverse direction. To retrieve the effective parameters the lattice constant is $3a/2$, which again is a signature of the double periodicity of the medium along the propagation direction. The effective index of refraction extracted from the transmission and reflection now displays a negative band that is associated with the double negativity, since both the effective compressibility and the effective mass density are negative in this frequency range. When adding the value of the experimental losses to the simulated cans, this negative band remains. Surprisingly, the losses broaden the negative band, which is in agreement with what we experimentally measured. We further observe that the imaginary part of the effective index of refraction reaches a value of approximately 1.5 in the negative band, which is not critical for observing transmission through a slab, as we showed experimentally. These effective parameters confirm the results from the band diagram and surface plot within the first Brillouin zone that we presented in Fig. 3.

Nevertheless, we chose not to present these results in the main text since we do not think that the description with effective parameters is the most relevant one. The main point of the main text is that breaking the symmetry of a single negative medium leads to the existence of a negative band. Even if the transmission through the two layers of cans evidences this band, most of the physics is hidden inside this transmission coefficient since it is the result of multiple scattering at the microscopic level within the unit cell. In the metamaterial community, it is commonly admitted that all scatterers see the same incident wave field, since they are arranged on a subwavelength scale, and thus the transmission is only a superposition of many resonators, all in phase. This transmission through two cans, which permits the negative band, clearly highlights that there is actually multiple scattering occurring at the microscopic scale of the medium, and that the two resonators do not see the same incident wave field.

Beam transmission through a slab demonstrates the negative refraction. In order to show that a band having a negative slope as shown in Fig. 3 is actually the signature of a medium exhibiting negative refraction, and that the super-resolution is not a canalization effect^{33,34}, we perform simulations on a slab geometry. We did not implement it experimentally because it would have required a much larger sample than the dimensions of our room. To avoid a too high number of mesh cells, those simulations only consist of simulating a slab layer of height $\sqrt{3}a$, which corresponds to the vertical unit cell of our honeycomb arrangement of soda cans. Periodic boundary conditions are then applied on the vertical boundaries to simulate the infinitely extended slab. This slab unit cell is excited by an incident plane wave with an incident angle θ with respect to the normal of the slab, and the phase shift applied on the periodic boundary conditions matches this value. We perform a set of 61 simulations with θ ranging from -90° to 90° . From those simulations, we are then able to build an incident wave field impinging on the slab that corresponds to a Gaussian beam with an incident angle θ_0 . To do so, we perform the complex summation of the wave fields extracted from each simulation and multiplied by a Gaussian coefficient $\exp\left(-\frac{(\theta-\theta_0)^2}{2\sigma_\theta^2}\right)$, where σ_θ corresponds to the angular aperture of the beam and is chosen to be equal to 9° . The complete field map on positions that are not within the simulation area is built by the use of the Bloch theorem for each incident angle.

Such a superposition of simulations is realized for distinct frequencies in the negative band (Extended Data Figs 4–6). We start by analysing the results corresponding to the frequency presented in the main text, that is, 417.5 Hz (Extended Data Fig. 4). For this frequency, the index of refraction is supposed to be -3 and we want to verify that the refraction satisfies this condition. From the simulations corresponding to distinct beam incident angles θ_0 , we evaluate the spatial Fourier transform of the field inside the crystal and isolate the dominating wavenumber in the first Brillouin zone. This permits us to find the Bloch wavenumber as well as the effective index of refraction inside the crystal. For the four angles represented in Extended Data Fig. 4, a value of -3 was confirmed. The ray tracing superimposed on the beams corresponds to this value, and is in very good agreement with the propagation of the beam through the slab as well as with the simulation of an effective medium (index of refraction of -3 and an effective density of -1.75 kg m^{-3}) represented on each subpanel. This simulation first confirms that we have negative refraction, since the outgoing beam seems to come from a y coordinate that is higher than the coordinate of the incident beam. Second, since the same effective index of refraction has been used for the four angles, this confirms that the refraction law can be described by an effective index of refraction instead of anisotropic parameters. Third, these results do not show any diffraction order at the exit of the slab as one would expect from periodic media. This comes from the fact that the medium is organized on a scale much smaller than the free-space wavelength and therefore all of the diffraction orders are evanescent at the exit interface, again confirming the description of refraction with a negative index. Nevertheless, these simulations also show that the reflection at the first interface is important, a consequence of the impedance mismatch. Some differences can be noted between the honeycomb medium and the effective one in the amplitude of the reflection, which means that the effective parameters that we extracted from the normal incidence on a single unit cell do not completely match the physics of the system. Probably more elaborate techniques that simulate the transmission with non-normal incidence on the unit cell would permit better effective parameters to be retrieved, but we have not found such sophisticated procedures in the literature. This is a reason why we opted in the main text not to describe the propagation in terms of the effective parameters, but rather insist on the fact that the double periodicity of the honeycomb lattice creates the negative refraction effect.

To give a more complete picture of the honeycomb soda-can metamaterial, we run other sets of simulations for different frequencies. In Extended Data Figs 5 and 6, we show the same kind of results for respective frequencies of 418.5 Hz and 419.5 Hz. The corresponding effective indices of refraction are respectively -2.4 and -1.5 , and the effective mass densities are respectively -1.35 kg m^{-3}

and -0.5 kg m^{-3} . Those simulations confirm the negative refraction effect previously observed, but some small differences can be noted. For example, for the incident angles of 30° and 60° the outgoing beam does not completely match the ray tracing. We attribute this effect to the existence of Fabry–Perot resonances within the slab's thickness, and the multiple reflections inside the slab vertically shift the outgoing beam. These Fabry–Perot resonances come from the impedance mismatch at the two interfaces of the slab, which indicates that the reflection is not null. Also, when looking at the map of the 60° incident beam (it is also visible on the others), one can see that the metamaterial slab creates backward reflection in the direction of the incident beam. Indeed, the intensity map shows cancellations and maxima of the field, which are the signature of stationary waves. The reflected beam (positive values of y) does not show these stationary effects. To our knowledge, this has not been described in the literature and would merit some deeper investigation.

Acoustic experiments. Experimental setup. The soda cans are placed in a honeycomb compact arrangement, with the apertures of all cans oriented in the same direction. Some acoustic absorber is first put on the short edges of the slab to prevent reflections and the formation of undesired edge states in the slab and second on walls surrounding the slab, to avoid reflections from the walls of the room. A two-dimensional translational stage with a 1-m range in each direction is used to scan the pressure field. A set of two microphones mounted on a perpendicular arm on the stage are used in order to increase the scanning area. The microphones are placed approximately one centimetre above the cans' apertures and are set to record sound with the same gain using an amplifier. An 8-cm-wide loudspeaker placed approximately 5 cm away from the slab's interface, in front of the middle can, produces a 2.5-s-long chirp signal, ranging from 200 Hz to 700 Hz. The two-sources experiment is performed by linear combination: a second one-source experiment is performed with the loudspeaker shifted by a distance $\lambda_0/7$ (in front of the first neighbouring can) and the maps are eventually obtained by summing the results of the two experiments (with a change of sign for the second one). The field maps are obtained by performing Fourier transforms of the recorded time dependent signals and are represented saturated. The loss-compensated maps are obtained by evaluating the field amplitude exponential decrease coefficient α on the line $y = 0$, that is, due to dissipative losses within the cans. The field within the slab at each x is then amplified by a factor $e^{\alpha x}$, with $\alpha = 0.0079 \text{ mm}^{-1}$ (equivalent to an attenuation length $L = 127 \text{ mm}$).

Extracting the imaginary part of the index of refraction. It is well known that the initial proposal of a perfect lens with a negative index of refraction is hampered by the presence of losses, which are inherent to the use of resonant unit cells to build the metamaterial. Even our Helmholtz resonators, that is, the soda cans, which have been chosen because they minimize losses, present some attenuation effects and hence cannot be considered as perfect resonators. In order to estimate those losses within the propagating band with a negative index of refraction, we evaluate the attenuation length in the corresponding frequency band. To do so, we average the absolute value of the complex field within the medium for each x coordinate. We find exponentially decreasing behaviour for each frequency, and we fit the logarithm of those curves with a linear fit. The slope $\alpha(\omega)$ corresponds to the inverse of the attenuation length, and the imaginary part of the index of refraction corresponds to $n''(\omega) = (c_0/\omega)\alpha(\omega)$.

In the results that are presented in Extended Data Fig. 3, we note that the attenuation is relatively important and is maximum for frequencies where the index of refraction takes the higher absolute value. These results are in agreement with the imaginary part of the effective index of refraction extracted from the simulation of the transmission through a unit cell of the metamaterial (Extended Data Fig. 2). Given the thickness of our slab in terms of the free-space wavelength, these losses are not critical and a sufficient part of the energy is transmitted through the slab to create a focus. This attenuation explains why the medium with a -1 index of refraction does not give perfect lensing, since the attenuation prevents the amplification of the evanescent part of the spectrum. The focus remains then diffraction limited for this index of refraction. Nevertheless, exploiting the high absolute values of the index of refraction at 417.5 Hz, as is done in the main text, allows us to observe the subwavelength spots even in the presence of attenuation, as is the case in our experimental setup.

Effective index extraction from tracking the focus. In the main text we only display the acoustic field maps corresponding to the focus beyond the diffraction limit at 417.5 Hz (Fig. 4). But the experimental data gives access to a broad frequency range. Since the negative branch is strongly dispersive, as shown by the band calculation in Fig. 3, the effective index varies. Here we show that tracking the position of the focus on the other side of the lens allows the measurement of this change in the effective index that characterizes the honeycomb arrangement of soda cans. To do so, we renormalize the acoustic intensity measured at different frequencies by the maximum of the intensity measured in the area behind the flat lens. This enhances the intensity in the focusing area and permits the focused wave

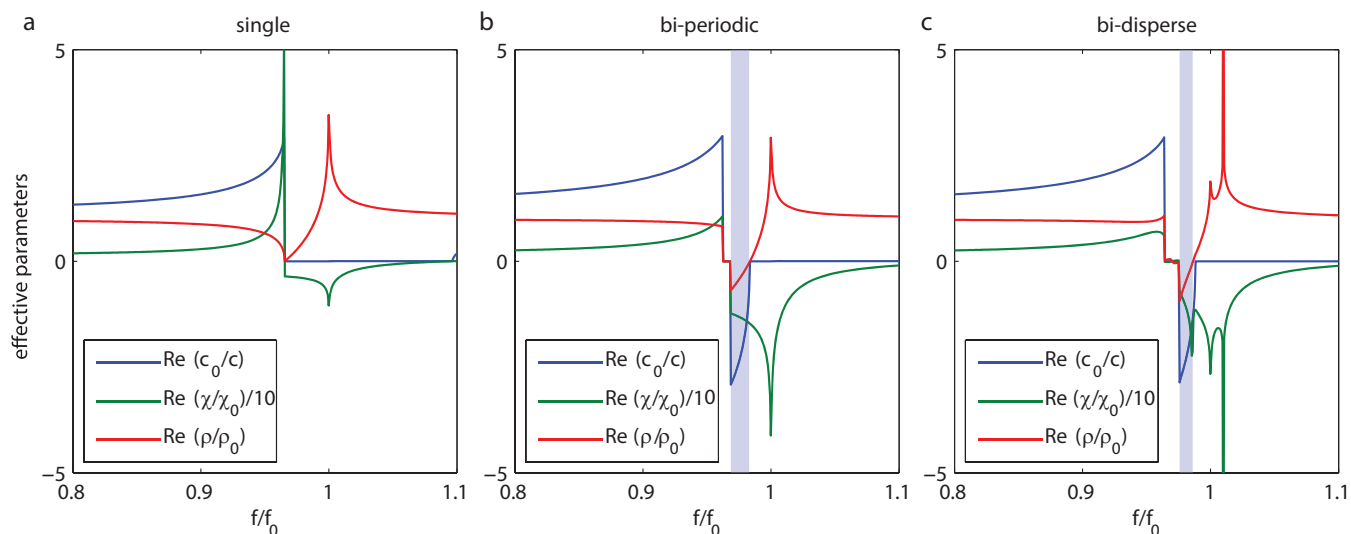
field to be observed. The result of such an operation is represented in Extended Data Fig. 7 for three different frequencies. The white dashed lines on the figures highlight the position of the focus along the depth (x direction). Depending on the index of refraction the focus is either stuck to the surface or well-detached, which can be easily explained by a ray tracing. For the three different maps we superimpose the ray tracing corresponding to an incident angle of 60° . To find the right index of refraction within the lens (between the two red dashed lines) we pick the value that permits a crossing at the focus to be seen. By doing so, we find respectively an index of refraction of -3 , -1.1 and -1 for the frequencies of 417, 427 and 445 Hz. The map at 417 Hz deserves one more comment: since the evanescent field is really strong, the focus seems to remain at the last layer of the cans while the maximum seems to be on top of the last can's aperture. But owing to the symmetry of the field while the losses are compensated within the lens (Fig. 4c) we chose the index of refraction that has a crossing inside the lens at half its thickness. The two other maps clearly show that the description with a negative effective index of refraction is totally relevant to predict the wave field transmitted through the lens, even if the wave field inside the experimental lens exhibits some hot spots due to the cans' apertures. Contrary to a perfect lens⁷, the focus shown on the last map where the index of refraction is -1 is diffraction limited. This comes from the inherent losses of the medium under study, characterized in a previous paragraph of the Methods.

The same procedure has been applied for all frequencies ranging from 415 to 450 Hz to track the focus position more systematically. In Extended Data Fig. 8 the profiles of the focus along the longitudinal direction (x direction in the previous maps) for all frequencies are displayed. From those profiles we clearly see the existence of two distinct regimes: a first one where the field is dominated by the evanescent waves and in which the focus seems to be attached to the lens and appeared at the position of the last layer of cans; and a second one in which the focus is well detached and it becomes feasible to find an effective index of refraction by ray tracing. In this second regime, we notice that the focus is moving further while the frequency is increased, meaning that the absolute value of the index of refraction is also decreasing. This is in agreement with

the band diagram presented in Fig. 3 but those experimental results surprisingly show that the negative band is broader than expected from the lossless simulation of the unit cell (Fig. 3b).

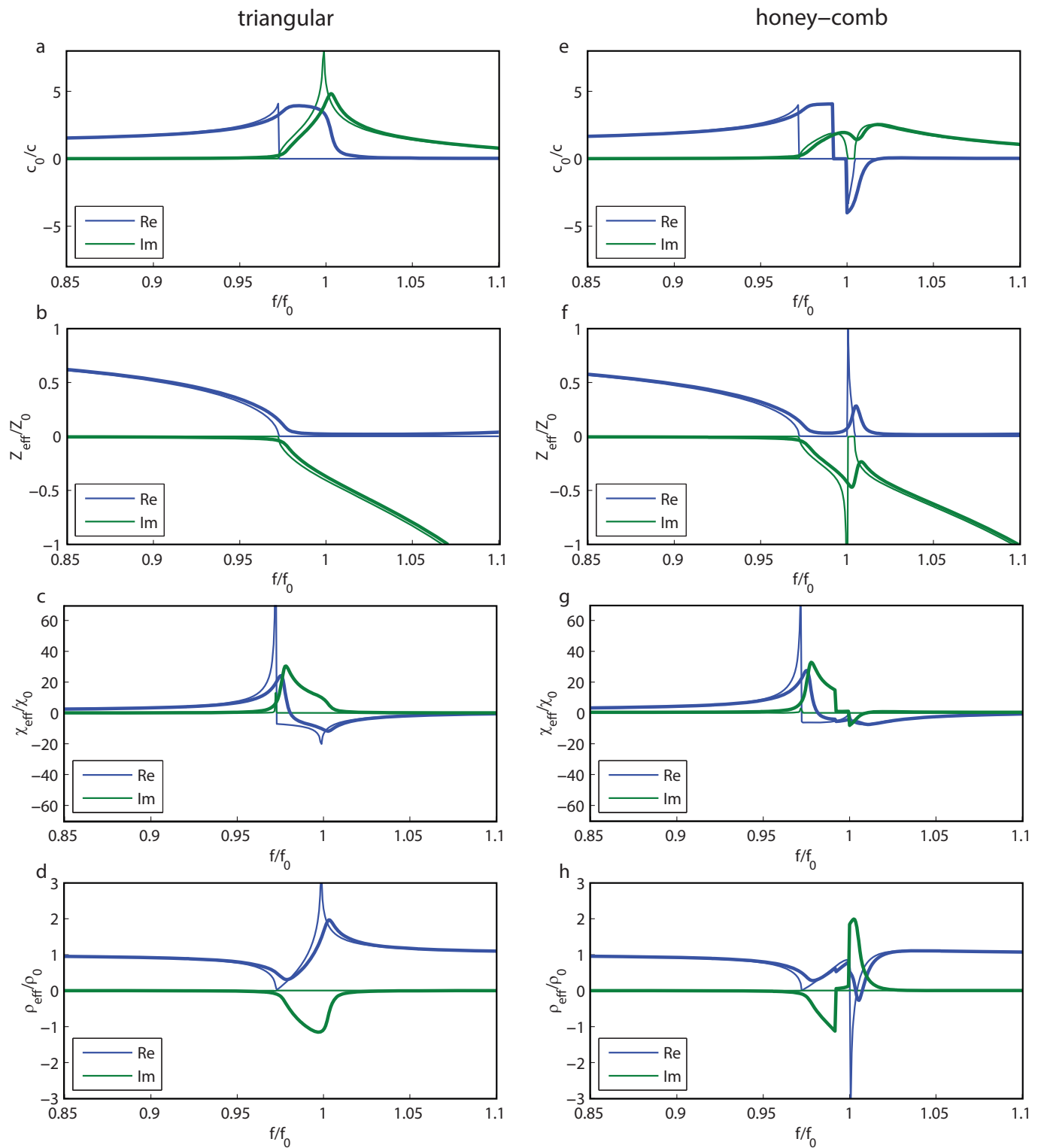
Changing the phase difference between the two sources. In the acoustic experiment presented in Fig. 4, we show that the field generated by two sources separated by $\lambda_0/7$ and emitting out of phase is reconstructed on the other interface of the superlens. This configuration may seem the optimal one since interferences created by the two sources generate by symmetry a zero of the field between them. In order to avoid any misinterpretation, we build the field maps obtained for any phase difference ranging from 0 to 360° between the two sources. From those maps we extract the intensity received in the focal plane. Then we build in Extended Data Fig. 9a a map of those intensity profiles as a function of the phase difference between the two sources. This map shows that whatever the phase shift the superposition of the two focal spots leads to the creation of two well separated foci. For the sake of completeness, we also represent in Extended Data Fig. 9b the real part of the field measured in the focal plane. The focused field presents the same dependence as the emitted one: the real part of the field measured at the focus corresponding to the first source varies from 1 to -1 and back to 1 while the phase difference between the emitting sources varies from 0 to 360° . From the map of the intensity one can notice that when the two sources are in quadrature the two intensities measured on the two focal spots are not strictly the same due to the existence of side lobes for a single focus. The two sources remain however separable in any configuration. As claimed in the main text, those results confirm that the resolution of the superlens is effectively $\lambda_0/7$.

31. de Vries, P., Van Coevorden, D. & Lagendijk, A. Point scatterers for classical waves. *Rev. Mod. Phys.* **70**, 447–466 (1998).
32. Fokin, V., Ambati, M., Sun, C. & Zhang, X. Method for retrieving effective properties of locally resonant acoustic metamaterials. *Phys. Rev. B* **76**, 144302 (2007).
33. Belov, P. A., Simovski, C. R. & Ikonen, P. Canalization of subwavelength images by electromagnetic crystals. *Phys. Rev. B* **71**, 193105 (2005).
34. Zhu, J. *et al.* A holey-structured metamaterial for acoustic deep-subwavelength imaging. *Nature Phys.* **7**, 52–55 (2010).



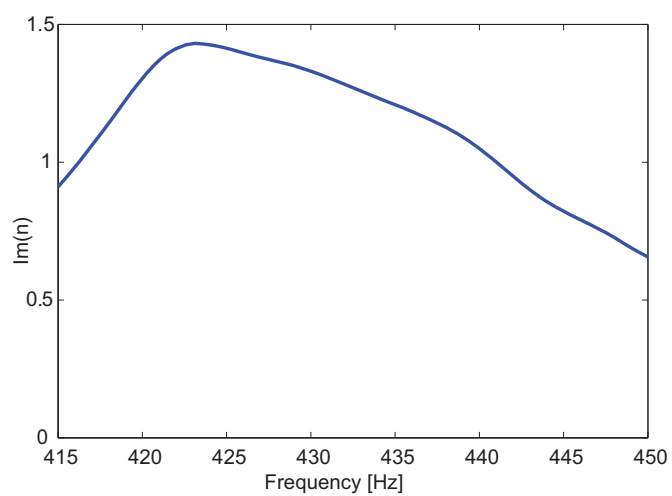
Extended Data Figure 1 | Effective parameters for metamaterials made from resonant point scatterers. **a**, Parameters for the single negative medium made of a periodic arrangement of identical point scatterers. **b**, **c**, Parameters

for the bi-periodic medium (**b**) and the bi-disperse medium (**c**), showing the double negativity for each medium. f , Frequency; f_0 , resonant frequency; see Methods for definitions of variables shown in the keys.

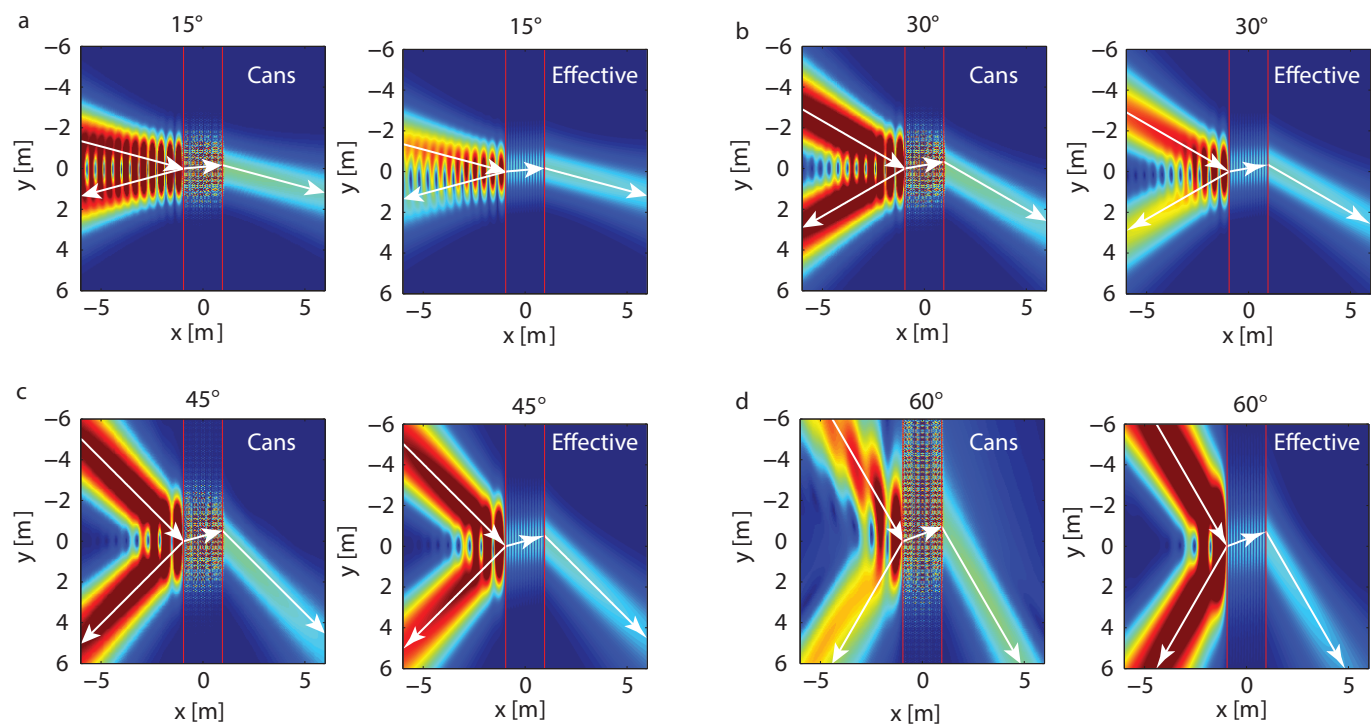


Extended Data Figure 2 | Effective parameters for the simulated soda-can metamaterials extracted from the transmission and reflection of a single unit cell. a–h, Speed of sound (a, e), impedance (b, f), compressibility (c, g) and

modulus (d, h) for the triangular (left column) and the honeycomb (right column) arrangements. Thick lines, with loss; thin lines, without loss; blue lines, real part; green lines, imaginary part.



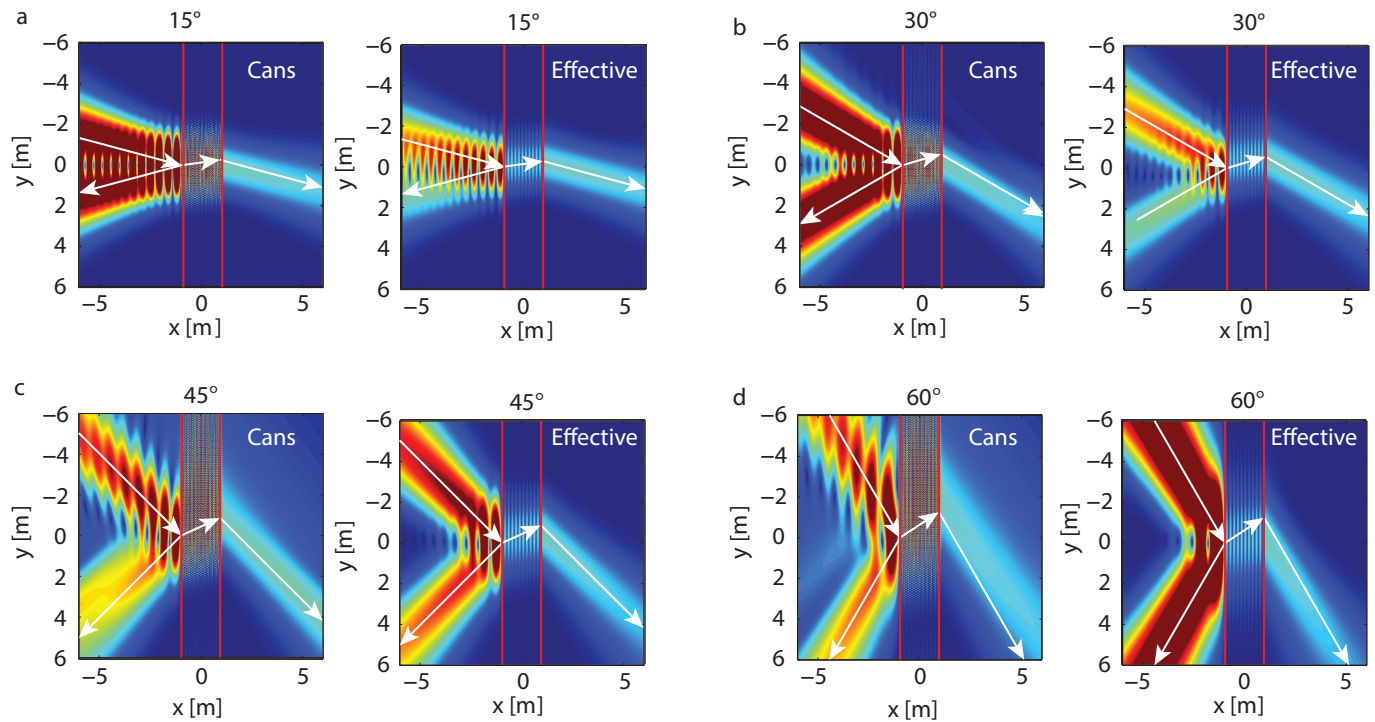
Extended Data Figure 3 | Experimental estimation of the imaginary part of the effective index of refraction within the negative band.



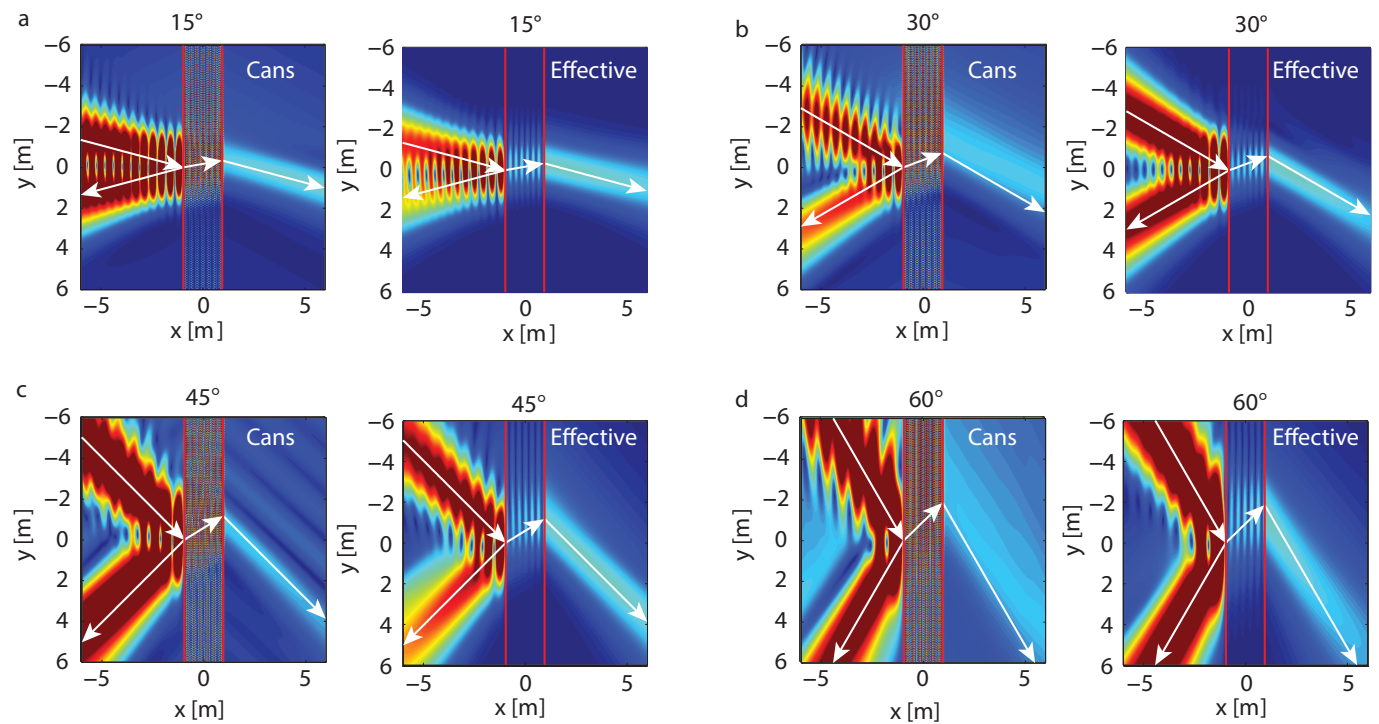
Extended Data Figure 4 | Simulated Gaussian beam incident on an infinite slab showing negative refraction at a frequency of 417.5 Hz.

a–d, Simulations for incident angles of 15° (**a**), 30° (**b**), 45° (**c**) and 60° (**d**); for each angle, data are shown for a lossless slab made of soda cans (left) or of

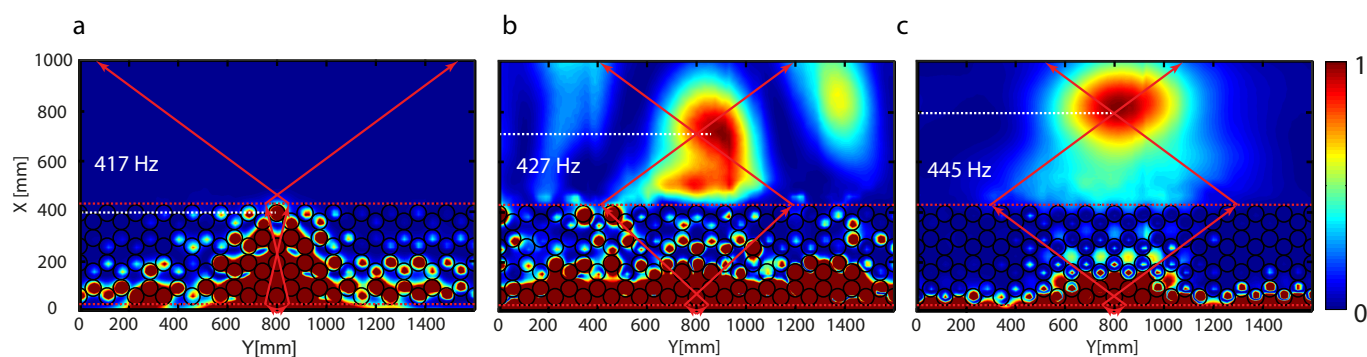
negative effective medium (right). The frequency corresponds to 417.5 Hz, for which the effective index of refraction is -3 . Red lines delimit the slab interfaces while white arrows evidence the direction of the incident, reflected and refracted beams.



Extended Data Figure 5 | Simulated Gaussian beam incident on an infinite slab showing negative refraction at a frequency of 418.5 Hz. As Extended Data Fig. 4 but for 418.5 Hz, and an effective index of refraction of -2.4 .

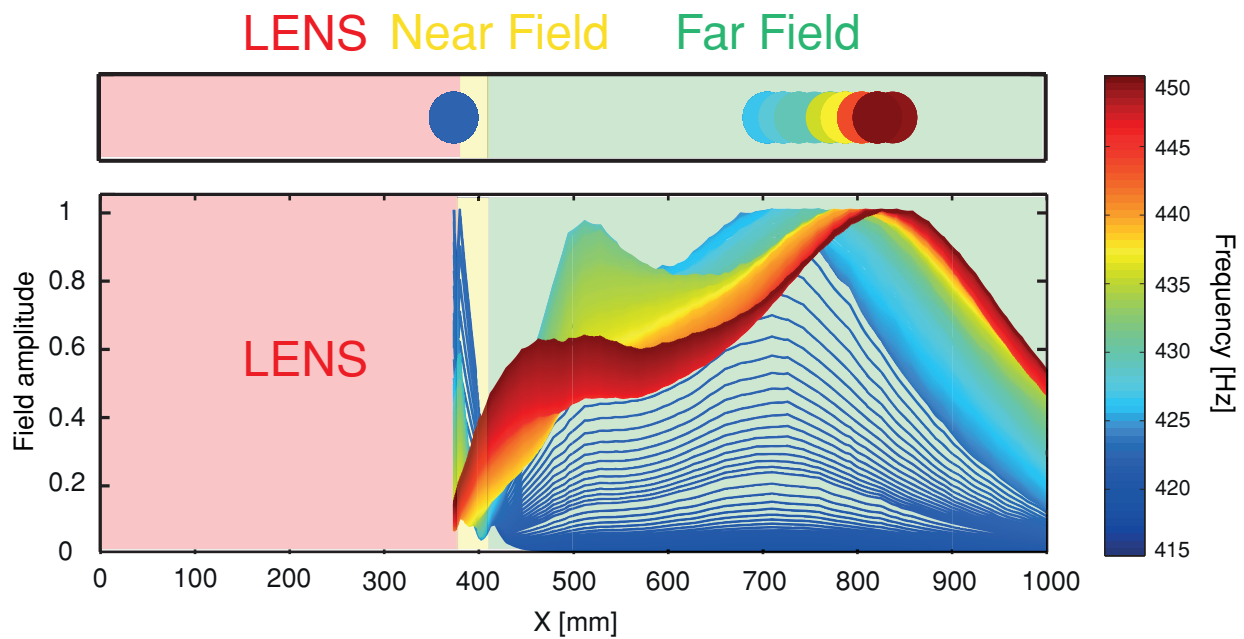


Extended Data Figure 6 | Simulated Gaussian beam incident on an infinite slab showing negative refraction at a frequency of 419.5 Hz. As Extended Data Fig. 4 but for 419.5 Hz, and an effective index of refraction of -1.5 .



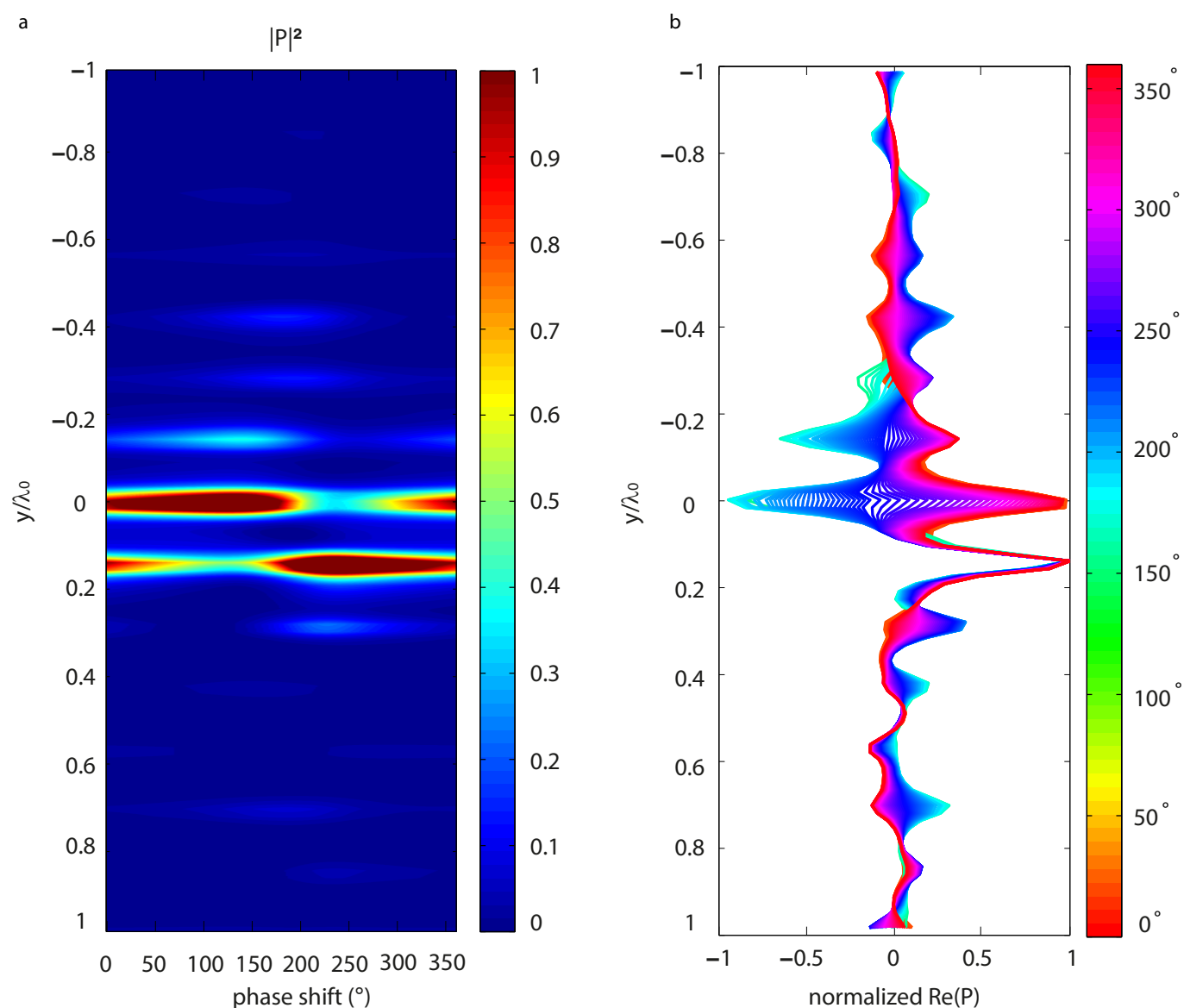
Extended Data Figure 7 | Renormalized maps of the measured acoustic pressure intensity for three different frequencies within the negative index of refraction band. a, $f = 417$ Hz; b, $f = 427$ Hz; and c, $f = 445$ Hz. For each

frequency the ray tracing is superimposed (red). The red dashed lines correspond to the slab's interfaces; the white dashed lines highlight the focus depth.



Extended Data Figure 8 | Longitudinal profiles of the focus for varying frequencies. Bottom panel: profiles of the field within the lens (red shading), in its near field (yellow shading) and its far field (green shading) for different

frequencies within the negative band (colour key). Top panel: tracked position of the foci in the previously described regions for the same frequencies in the negative band (colour key).



Extended Data Figure 9 | Effect of the phase shift between the two sources on the intensity measured on the focal plane. **a**, Map of the intensity of the field in the focal plane along the normalized axis y/λ_0 (colour coded, key at

right) as a function of the phase shift between the sources. **b**, Profiles of the normalized real part of the field ($\text{Re}(P)$) along the normalized axis y/λ_0 for varying phase shifts (colour key) between the two sources.

Guiding the folding pathway of DNA origami

Katherine E. Dunn^{1†*}, Frits Dannenberg^{1,2*}, Thomas E. Ouldridge^{3†}, Marta Kwiatkowska², Andrew J. Turberfield¹ & Jonathan Bath¹

DNA origami is a robust assembly technique that folds a single-stranded DNA template into a target structure by annealing it with hundreds of short ‘staple’ strands^{1–4}. Its guiding design principle is that the target structure is the single most stable configuration⁵. The folding transition is cooperative^{4,6,7} and, as in the case of proteins, is governed by information encoded in the polymer sequence^{8–11}. A typical origami folds primarily into the desired shape, but misfolded structures can kinetically trap the system and reduce the yield². Although adjusting assembly conditions^{2,12} or following empirical design rules^{12,13} can improve yield, well-folded origami often need to be separated from misfolded structures^{2,3,14–16}. The problem could in principle be avoided if assembly pathway and kinetics were fully understood and then rationally optimized. To this end, here we present a DNA origami system with the unusual property of being able to form a small set of distinguishable and well-folded shapes that represent discrete and approximately degenerate energy minima in a vast folding landscape, thus allowing us to probe the assembly process. The obtained high yield of well-folded origami structures confirms the existence of efficient folding pathways, while the shape distribution provides information about individual trajectories through the folding landscape. We find that, similarly to protein folding, the assembly of DNA origami is highly cooperative; that reversible bond formation is important in recovering from transient misfoldings; and that the early formation of long-range connections can very effectively enforce particular folds. We use these insights to inform the design of the system so as to steer assembly towards desired structures. Expanding the rational design process to include the assembly pathway should thus enable more reproducible synthesis, particularly when targeting more complex structures. We anticipate that this expansion will be essential if DNA origami is to continue its rapid development^{1–3,17–19} and become a reliable manufacturing technology²⁰.

This study is based on a simplified version of the archetypal origami tile¹ and, in particular, on the distribution of observed folds of a ‘dimer’ variant which contains two copies of the template sequence in head-to-tail repeat. The ‘monomer’ tile (Fig. 1) is created by annealing a 2,646-nucleotide (nt) circular template with 90 staples, each designed to hybridize to one or more 15- or 16-base domains of the template. 76 of the staples mediate interactions between pairs of non-contiguous template domains, as follows: 66 U-shaped ‘body’ staples form short-range contacts between domains that are relatively close in the primary sequence of the template; and 5 pairs of ‘seam’ staples form long-range contacts, bridging between positions where the template folds back on itself to form a central seam¹. Unlike the interactions between amino acid residues that stabilize a protein, staples mediate interactions between template domains that are highly specific: each staple can be considered to bind stably only to complementary domains of the template. The designed fold of the monomer tile corresponds to an absolute minimum in the free energy landscape. This origami folds with high yield to form discrete rectangular tiles

of approximately 80 nm × 40 nm (Fig. 1c); approximately 80% of tiles appear to be well folded.

The ‘dimer’ template is also circular. It contains two identical copies of the monomer joined head-to-tail and can therefore bind two copies of each staple (Fig. 2). Each pair of body and seam staples can bind in one of two configurations (Fig. 2a) to form either an internal link within each copy of the monomer sequence or a pair of cross-links between the two copies. The total number of possible domain pairings is $2^{76} \approx 10^{23}$. Although many of these configurations are sterically inaccessible, it is clear that the result of reducing the specificity of staple binding is that, as in the case of protein folding, the number of possible states of the system is overwhelmingly greater than the number of well-folded structures. However, in contrast to proteins (and to conventional origami structures) there is more than one ‘well-folded’ state (Fig. 2): not one but a handful of well-folded states occupy discrete energy minima in a vast configurational landscape. Remarkably, when the dimer origami is annealed by cooling from 95 °C, a small set of well-folded shapes are formed with good yield: each consists of a pair of rectangular tiles attached on one edge (Fig. 2b, c). The probability of finding well-folded structures by random search of configuration space is negligible⁸, therefore efficient folding pathways must exist^{21,22}. As in protein folding, assembly is constrained such that the system is highly likely to discover free-energy minima that correspond to well-formed final states.

The dimer origami tile has 22 template routings that correspond to well-folded configurations in which all staple binding sites are occupied and in which the tile is expected to be planar and unstrained. These give 6 unique shapes, each with a characteristic offset between two linked rectangular components which have essentially the same structure as the monomer tile (Fig. 3a–c and Extended Data Fig. 1). These shapes can be grouped into classes according to the contacts made by the seam staples: fold $m:n$ has m pairs of seam staples that

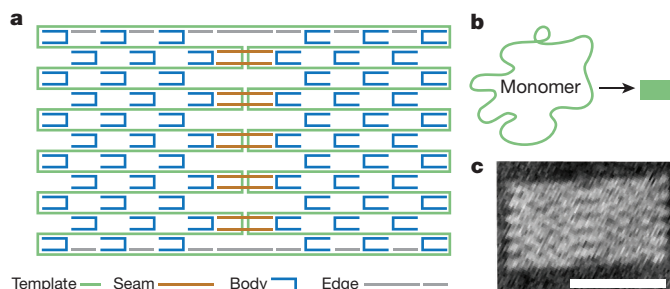


Figure 1 | The monomer tile. **a**, 66 body staples (blue) and 5 pairs of seam staples (brown) each hybridize to two non-contiguous domains of the circular template. Edge staples (grey) fill gaps at the top and bottom of the structure. Hybridization of body and seam staples pins the corresponding domains of the template together, determining the unique stable, rectangular fold of this simplified origami tile as indicated in **b**. **c**, Atomic force micrograph of the monomer tile (scale bar, 50 nm).

¹University of Oxford, Department of Physics, Clarendon Laboratory, Parks Road, Oxford OX1 3PU, UK. ²University of Oxford, Department of Computer Science, Wolfson Building, Parks Road, Oxford OX1 3QD, UK. ³University of Oxford, Department of Physics, Rudolf Peierls Centre for Theoretical Physics, 1 Keble Road, Oxford OX1 3NP, UK. [†]Present addresses: Department of Electronics, University of York, York YO10 5DD, UK (K.E.D.); Department of Mathematics, Imperial College, London SW7 2AZ, UK (T.E.O.).

*These authors contributed equally to this work.

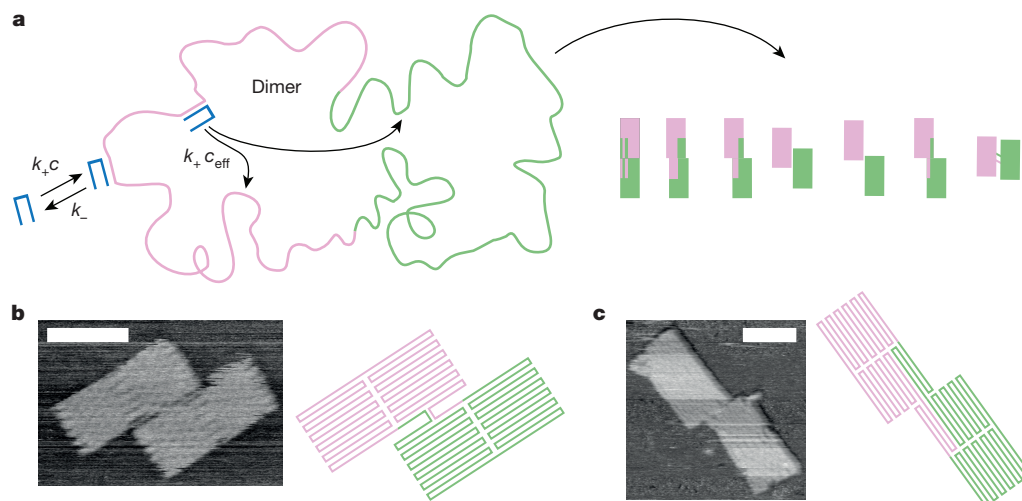


Figure 2 | Folding origami tiles with a dimer template. **a**, Left, the base sequence of the green section of the template is the same as that of the pink section, so the dimer template can fully hybridize to two copies of each staple. Arrows indicating staple binding and unbinding transitions are annotated with reaction rates used in a model of assembly. Two identical staples can bind to the template in one of two configurations, binding together pairs of domains within each section or connecting domains in different sections. This gives a

vast number of possible configurations, a handful of which are well-folded (shown right, and defined in Fig. 3 and Extended Data Figs 1 and 2). Ordered folds of the dimer template comprise two linked rectangular tiles with a characteristic offset on the long or short edges, for example, **b** or **c**, respectively (in each case, an AFM image with a scale bar of 50 nm is shown alongside a schematic of the fold).

connect domains within each half of the template and n pairs of seam staples that form connections between domains in opposite halves (Fig. 3a–c). Folds $m:n$ and $n:m$ are related by symmetry and are therefore not distinguished in our experiments or analysis (Extended Data Fig. 1b). A set of non-planar folds adds a seventh shape to the six defined above and a further 52 template routings (Fig. 3d and Extended Data Fig. 2). Fink and Ball²³ have estimated the maximum number of distinct, compact configurations that can be encoded into a single polymer sequence: for a polymer of 168 unique domain types on a square lattice^{23,24} the theoretical limit is 13. A major factor in allowing the large number of folds in our system is the extensive re-use of structural motifs within distinct folds, a possibility not considered by Fink and Ball.

Atomic force microscopy (AFM) enables us to distinguish different configurations of the template and this provides a unique opportunity to study folding pathways. Samples of annealed origami were imaged by AFM. Most observed shapes are consistent with the classification scheme shown in Fig. 3, and the outlines of 44% of objects identified as candidate dimer tiles were successfully fitted to measure the offset between the two component monomer tiles (Fig. 3e and Extended Data Fig. 3).

The distribution of tile shapes was compared to predictions made using a Markov chain model of folding in which each transition corresponds to binding or unbinding of a single staple domain (Fig. 2, Methods section ‘Folding model’). An unbound staple at concentration c binds to the template with a rate $k_+ c$ (where $k_+ = 10^6 \text{ M}^{-1} \text{ s}^{-1}$, ref. 25). After one half of a staple has bound, the second half can bind with a rate $(k_+ c_{\text{eff}})$ that depends on its effective concentration, c_{eff} , at the corresponding template domain. The effective concentration depends on the proximity of the template domain which, in turn, depends on the contacts between template domains already established by hybridization of other staples. We expect folding to be dominated by short-range interactions because staples are more likely to connect two template domains that are spatially close, either because they are closely spaced along the template or because the previous binding of other staples is holding them together. To determine the effective concentration, the shortest path through the part-assembled origami that connects the complementary template and staple domains is identified. This connection is modelled as a heterogeneous freely jointed chain with double-stranded (ds) and single-stranded (ss)

DNA components. The effective concentration of the part-bound staple at the complementary template domain is related to the probability that the ends of the chain lie spontaneously within a (short) interaction range. Unbinding of a staple domain is treated as a two-state transition, with a configuration-independent rate: $k_- = k_+ \exp\{\Delta G^{0,\text{duplex}}/RT\}(1 \text{ M})$ where $\Delta G^{0,\text{duplex}}$ is the change in standard free energy on forming the duplex at standard concentrations of 1 M. In order to represent steric constraints on folding, the state space of the model is restricted to patterns of staple binding in which each segment of the partially folded origami occurs in one of a set of pre-defined, well-ordered folds.

The histograms in Fig. 4 show distributions of offset values, measured by fitting AFM data (Extended Data Figs 3 and 4), and the corresponding distributions between the discrete shapes shown in Fig. 3 that are predicted by the model. Figure 4a corresponds to the staple set described above (see Fig. 1): structures with each of the seam configurations 5:0, 4:1 and 3:2 are observed. The model suggests that the folding pathway depends on competition between body and seam staples. If local interactions mediated by body staples were to form first and dominate the outcome, the system would prefer the 5:0i fold (see Fig. 3f for nomenclature) in which all body staples are bound to two domains that are as close as possible along the template. In this fold, no staples link the two halves of the template. However, strong seam connections that are inserted early in the folding pathway favour a more uniform distribution between all possible seam configurations: for example, once the part-folded structure 1:1 has formed, the 5:0 fold is inaccessible unless at least one seam connection is broken (Extended Data Fig. 5). With the staple set shown in Fig. 4a, each seam contact is bridged by two staples. The cooperative binding of seam staple pairs offsets the increased entropic cost of forming long-range contacts, with the result that seam staples are incorporated at a similar temperature to body staples in both model (Fig. 4a) and experiment (Extended Data Fig. 6). Consequently the model predicts that all seam configurations should be observed, consistent with experimental observations.

We predict that the folding pathway can be changed by altering the relative strengths of short- and long-range interactions. Breaking in half one of each pair of seam staples (Fig. 4b), so that the pairs no longer bind cooperatively, weakens these long-range bonds, causing them to form later in the folding pathway (Fig. 4b, central panel and Extended Data Fig. 6) and to break and reform in alternative configurations more frequently (Extended Data Fig. 7). With weakened

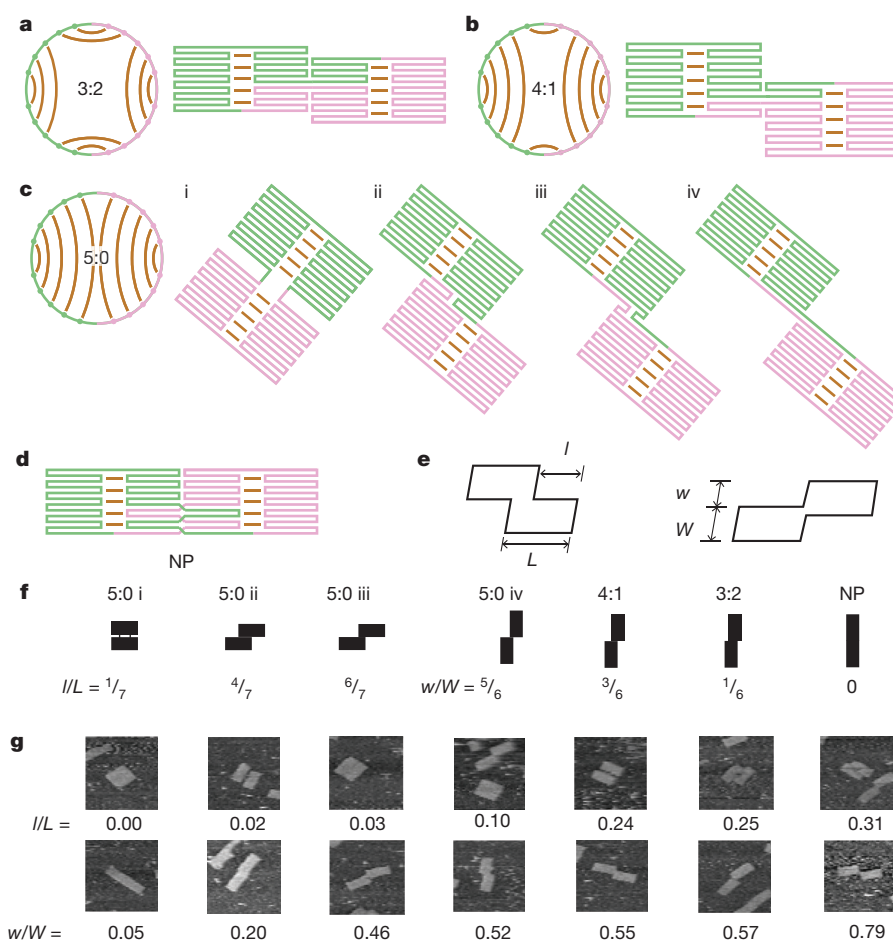


Figure 3 | Classification of well-folded shapes. **a–c**, Folds of the dimer origami tile can be classified by the pattern of interactions mediated by the seam staples: these contacts are shown schematically in diagrams in which the dimer template is represented as a circle. Fold $m:n$ has m seam contacts between domains within each copy of the monomer sequence and n seam contacts that form connections between the two copies (that is, connecting template domains of different colour). The fold 5:0 can be further divided into shapes that differ in the offset along the long edge of the two tiles (5:0i, 5:0ii and so on: **c**). **d**, The set of legal folds allowed by the model includes the configurations shown

in **a–c** and an additional set of non-planar configurations (NP), one of which is shown (see Extended Data Fig. 2 for the complete set). **e**, Tiles observed in folding experiments can be classified according to the fractional offset along the short or long edge of the tile (w/W and l/L respectively). **f**, Seven unique shapes corresponding to well-folded configurations (see also Extended Data Fig. 1). **g**, Gallery of shapes observed by AFM in a typical experiment with measured fractional offsets (each image is 300×300 nm). A bin size of 0.1 is used in histograms of fitted fractional offsets (Fig. 4).

long-range interactions, we expect folding to be governed primarily by local interactions. The model predicts that the distribution of shapes is shifted strongly towards the 5:0i fold, in which all body staples span the smallest possible distances along the template (the same distances as in the monomer tile), and this is confirmed by experiment (Fig. 4b). The thermodynamic cost of breaking every other seam staple is approximately equal in each well-folded state and therefore this change should not affect their equilibrium populations. We have changed the distribution between folds not by changing the relative stability of the final states but by deliberately controlling the stabilities of crucial intermediate states, thus shaping the folding pathway.

The importance of stable, long-range interactions in determining the folding pathway is revealed by the evolving correlations between seam staples in the model. Characteristic patterns of correlation can be used to predict the final fold even before seam staple occupancy has reached 50% (Extended Data Figs 8 and 9).

The influence of seam staples on folding is similar to that of disulphide bonds in Anfinsen's experiment on protein folding⁹. If long-range bonds are allowed to form first and, effectively, irreversibly, then folding is kinetically trapped. If they are weakened and permitted to rearrange then folding can be controlled by weaker short-range interactions.

Figure 4c shows an alternative staple set incorporating extended staples that form particularly strong short-range connections and therefore bind to the template early in the folding process (Fig. 4c, central panel). Without interference from other staples, these contacts are most likely to form between the pairs of template domains with the smallest separation along the template. These preferred contacts occur in the 3:2 and 5:0 folds but not the 4:1 fold (in the 4:1 fold, one extended staple forms a long-range contact between the two halves of the template). Experimental results confirm the model prediction that the 4:1 fold is strongly suppressed (Fig. 4c). As with the broken seam staples (Fig. 4b), this modification guides the folding pathway without imposing an energetic penalty on alternative folds.

We can control the fold of the dimer very effectively by engineering both the folding pathway and the stability of the chosen target structure. The 3:2 configuration can be favoured by weakening the original seams (as in Fig. 4b) and adding new seam staples that bridge between the monomer tiles without distortion only in the 3:2 configuration (Fig. 4d). This modification guides folding by increasing the stability of 3:2 relative to other folds. Similarly, a long staple in the bottom right corner of the monomer tile (Fig. 4e) biases folding towards the 5:0iv shape by decreasing the stability of other folds, which would require introduction of a sharp bend within the long staple. (The model does

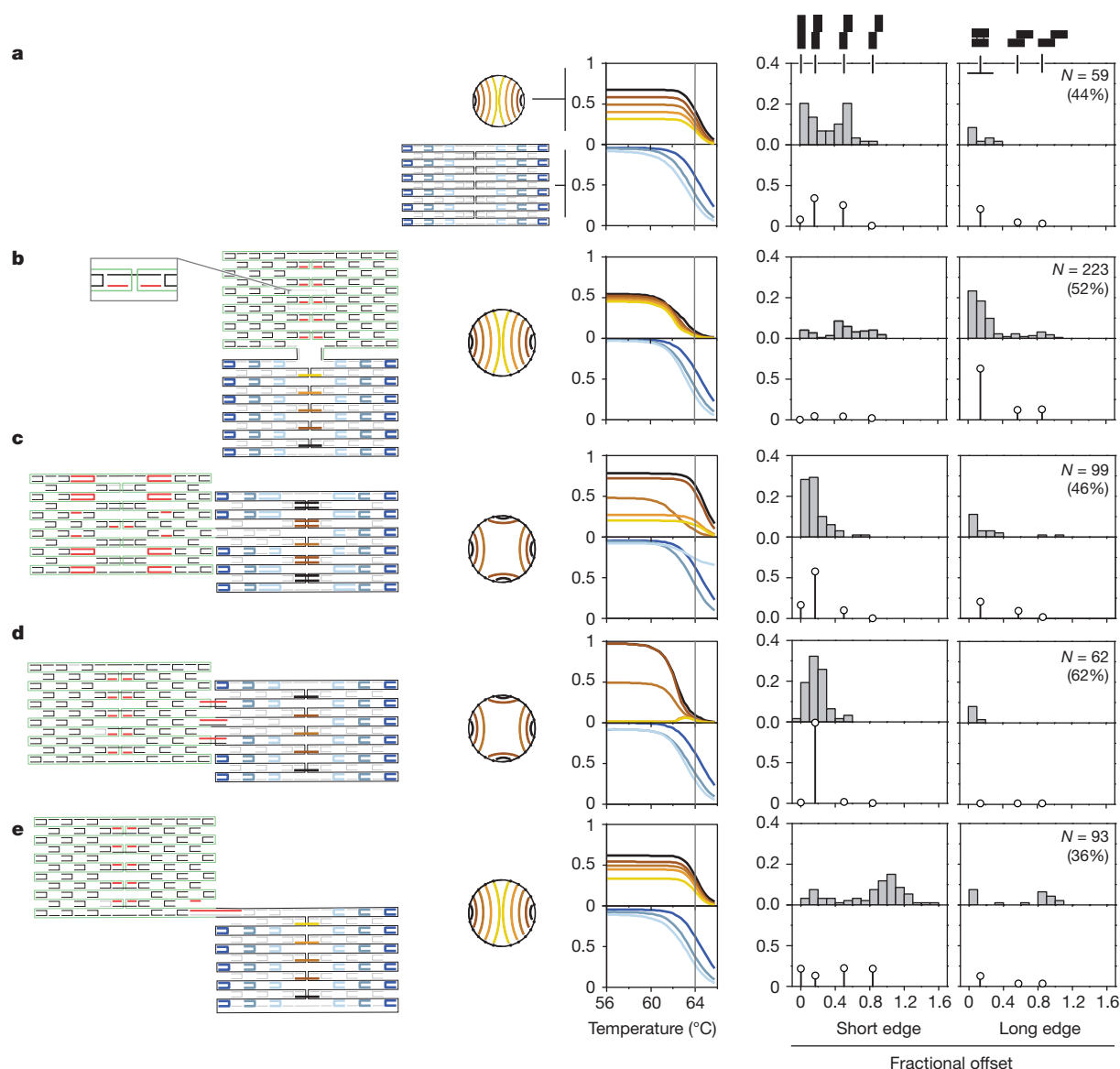


Figure 4 | Folding can be guided by modifying staples to steer the folding pathway. The reference staple set (a) folds to give a distribution of shapes that are characterized by the fractional offset between the two component tiles along the long or short edge. Modifications to the reference staple set (b–e) were designed to fold into specific target shapes. Left-hand panels show the staple configurations and the seam-staple contacts in the target structures. The top left rectangle of each target shape is used to highlight modified staples in red. The distance between the two template domains linked by a staple depends on fold: in the bottom right rectangle, staples are grouped and coloured according to the distance spanned (see key in a: short-range body staples are blue, seam staples are brown to yellow; lighter shades indicate larger distances).

not include any penalty for bending and so fails to predict the engineered bias in this case.)

By showing that an origami tile with a duplicated template can be annealed to produce a high yield of well-folded structures from among $\sim 10^{23}$ disordered alternative staple configurations, our results confirm that, as in the case of proteins, efficient folding pathways exist and that folding is highly cooperative. We infer that the folding of all DNA origami is shaped by similar pathways. Manipulation of the folding pathway validates our simple folding model, which successfully predicts the dominant folding pathways observed in experiments. We anticipate that this tool will prove more generally useful, to establish how to change the relative strengths of local and long-range staple interactions to rationally steer the folding pathway towards desired target structures.

Graphs in the central panels show the calculated fraction of contacts formed for each staple group (whether or not as part of the target structure) as a function of temperature during assembly. The right-hand panels show the distribution of shapes predicted and observed for each set: a histogram representing the continuous distribution of fitted offset values is plotted above the distribution between discrete shapes predicted by the model (indicated by silhouettes). The number of fitted shapes, N , and the yield (shapes fitted as a percentage of candidate structures identified) are shown in the top right corner of each histogram.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 24 October 2014; accepted 16 June 2015.

Published online 19 August 2015.

1. Rothmund, P. W. K. Folding DNA to create nanoscale shapes and patterns. *Nature* **440**, 297–302 (2006).
2. Douglas, S. M. *et al.* Self-assembly of DNA into nanoscale three-dimensional shapes. *Nature* **459**, 414–418 (2009).
3. Dietz, H., Douglas, S. M. & Shih, W. M. Folding DNA into twisted and curved nanoscale shapes. *Science* **325**, 725–730 (2009).
4. Sobczak, J. P. J., Martin, T. G., Gerling, T. & Dietz, H. Rapid folding of DNA into nanoscale shapes at constant temperature. *Science* **338**, 1458–1461 (2012).
5. Seeman, N. C. DNA in a material world. *Nature* **421**, 427–431 (2003).
6. Arbona, J.-M., Aimé, J.-P. & Elezgaray, J. Cooperativity in the annealing of DNA origamis. *J. Chem. Phys.* **138**, 015105 (2013).

7. Song, J. *et al.* Direct visualization of transient thermal response of a DNA origami. *J. Am. Chem. Soc.* **134**, 9844–9847 (2012).
8. Levinthal, C. How to fold graciously. Mössbauer spectroscopy in biological systems. *Univ. Illinois Bull.* **67**, 22–24 (1969).
9. Anfinsen, C. B. Principles that govern the folding of protein chains. *Science* **181**, 223–230 (1973).
10. Dobson, C. M. Protein folding and misfolding. *Nature* **426**, 884–890 (2003).
11. Baker, D. A surprising simplicity to protein folding. *Nature* **405**, 39–42 (2000).
12. Martin, T. G. & Dietz, H. Magnesium-free self-assembly of multi-layer DNA objects. *Nature Commun.* **3**, 1103 (2012).
13. Ke, Y., Bellot, G., Voigt, N. V., Fradkov, E. & Shih, W. M. Two design strategies for enhancement of multilayer-DNA-origami folding: underwinding for specific intercalator rescue and staple-break positioning. *Chem. Sci.* **3**, 2587–2597 (2012).
14. Castro, C. E. *et al.* A primer to scaffolded DNA origami. *Nature Methods* **8**, 221–229 (2011).
15. Douglas, S. M., Bachalet, I. & Church, G. M. A logic-gated nanorobot for targeted transport of molecular payloads. *Science* **335**, 831–834 (2012).
16. Perrault, S. D. & Shih, W. M. Virus-inspired membrane encapsulation of DNA nanostructures to achieve *in vivo* stability. *ACS Nano* **8**, 5132–5140 (2014).
17. Han, D. *et al.* DNA gridiron nanostructures based on four-arm junctions. *Science* **339**, 1412–1415 (2013).
18. Ke, Y., Ong, L. L., Shih, W. M. & Yin, P. Three-dimensional structures self-assembled from DNA bricks. *Science* **338**, 1177–1183 (2012).
19. He, Y. *et al.* Hierarchical self-assembly of DNA into symmetric supramolecular polyhedra. *Nature* **452**, 198–201 (2008).
20. Kershner, R. J. *et al.* Placement and orientation of individual DNA shapes on lithographically patterned surfaces. *Nature Nanotechnol.* **4**, 557–561 (2009).
21. Wolynes, P. G., Onuchic, J. N. & Thirumalai, D. Navigating the folding routes. *Science* **267**, 1619–1620 (1995).
22. Onuchic, J. N., Wolynes, P. G., Lutheyschulten, Z. & Socci, N. D. Towards an outline of the topography of a realistic protein-folding funnel. *Proc. Natl Acad. Sci. USA* **92**, 3626–3630 (1995).
23. Fink, T. M. A. & Ball, R. C. How many conformations can a protein remember? *Phys. Rev. Lett.* **87**, 198103 (2001).
24. Jacobsen, J. L. & Kondev, J. Field theory of compact polymers on a square lattice. *Nucl. Phys. B* **532**, 635–688 (1998).
25. Zhang, D. Y. & Winfree, E. Control of DNA strand displacement kinetics using toehold exchange. *J. Am. Chem. Soc.* **131**, 17303–17314 (2009).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank K.V. Gothelf, M. Dong, A.L.B. Kodal, S. Helmig and S. Zhang (Department of Chemistry and Interdisciplinary Nanoscience Centre iNano, Aarhus, Denmark) for assistance with AFM imaging. This research was supported by Engineering and Physical Sciences Research Council grants EP/G037930/1 and EP/P504287/1, a Human Frontier Science Program grant RGP0030/2013, a Microsoft Research PhD Scholarship (F.D.), the ERC Advanced Grant VERIWARE (F.D. and M.K.) and a Royal Society–Wolfson Research Merit Award (A.J.T.).

Author Contributions K.E.D. performed the experimental work, F.D. and T.E.O. developed the folding model, J.B. and A.J.T. devised the experimental strategy. All authors contributed to experimental design, interpretation of the data and preparation of the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.B. (j.bath@physics.ox.ac.uk) or A.J.T. (a.turberfield@physics.ox.ac.uk).

METHODS

Experimental methods. Plasmid pUC19 cut with HindIII and EcoRI was amplified by PCR with the primers TGACCTAATCCTCAGCAATTCAGTGGCC GTCGTTTTACAA and ACGGACGCGCTGAGGAGCTTGGCGTAATCATG GTCATAG in order to trim the template to the desired length and introduce a unique BbvCI site. The PCR product was cut with BbvCI and ligated to generate pKD1 (2,646 bp). A typical monomer plasmid preparation contains a small amount (~1%) of plasmid dimer. The dimer plasmid was obtained by nicking a monomer plasmid preparation with Nt.BbvCI (in order to resolve monomer and dimer more easily), purifying the nicked dimer band from a 0.7% TAE agarose gel, then transforming the purified nicked dimer into the *recA* host DH5 α . The template sequence is given in Supplementary Information.

Single-stranded template was prepared by sequential reaction of either monomer or dimer pKD1 with Nt.BspQI at 50 °C and ExoIII at 37 °C to digest the non-template strand and leave a covalently-closed single-stranded template²⁶. Enzymes were removed by phenol:chloroform extraction and the template was recovered by ethanol precipitation; its concentration was then determined by measuring ultraviolet absorbance at 260 nm.

DNA origami was designed using caDNA²⁷ and was assembled by cooling template at 4–10 nM with a ~10-fold excess of staples from 95 °C to 25 °C at 1 °C per minute in a buffer containing 40 mM Tris-acetate (pH 8.3) and 12.5 mM magnesium acetate. Excess staples were removed using an S-300 size exclusion spin column²⁸. Staple sequences for the standard design and variations are given in Supplementary Information.

Atomic force microscopy images were acquired using either an Agilent 5500 AFM with Olympus TR400-PSA probes (Figs 1–3, 4a) or a Veeco Dimension 3100 with Bruker SNL-10 probes (all other figures). A few microlitres of sample were added to freshly cleaved mica and the sample was imaged in tapping mode in an imaging buffer containing 12.5 mM magnesium acetate, 4 mM NiCl₂, 1 mM EDTA and 40 mM Tris-acetate pH 8.0–8.3 (the imaging buffer for Fig. 1c lacked NiCl₂, the imaging buffer for Fig. 2c lacked EDTA).

Folding model. Our domain-level description of origami assembly is intended to reproduce some aspects of cooperativity. In particular, it accounts for the increase in incorporation rate for a staple when its target domains on the template are held more closely together as a result of the earlier binding of other staples. This effect is most noticeable in the seam where the binding of the first of a pair of seam staples greatly accelerates, and is stabilized by, the binding of the second. The model incorporates a physically reasonable approximation of the entropic cost of closing loops by staple binding, but is far from a complete description of the physics of assembly. It is useful in guiding, and providing insights into, the effects of significant changes to the origami design.

We model the folding of an isolated template in the presence of an excess of staples as an inhomogeneous continuous-time Markov chain. Each transition between states corresponds to the binding or unbinding of a single staple domain. Transition rates between two states are chosen according to an estimate of the free energy difference between the two, in a manner that would reproduce the correct Boltzmann distribution if this free energy difference were calculated exactly. The temperature is updated once per second of simulated time which allows us to use an event-based Gillespie simulation algorithm²⁹ with transition rates fixed over one second intervals. Data on folding processes are collected by simulating multiple folding trajectories (typically 1,600 per experiment).

Subsequent sections contain more detailed descriptions of the folding model. **State space.** We consider the possible configurations of staples hybridized to the template with domain-level resolution: a domain is either fully hybridized or unhybridized. A staple is called half-bound if only one of its two domains is hybridized to the template and fully bound if both domains are bound. In the model, a staple domain can only hybridize to the complementary template domain; we ignore weaker interactions that result from inevitable partial sequence complementarity between other pairs of domains.

For each type of two-domain staple (and the corresponding two pairs of complementary template domains) there are 34 distinct patterns of domain binding (states) with between zero and four copies of the staple bound to the dimer template. One is an empty state. When one staple is bound to the template there are four states in which the staple is half-bound and four states in which the staple is fully bound. When two staples are bound to the template there are six states in which both staples are half-bound, eight states with one half-bound and one fully bound staple, and two states with two fully bound staples. There are four states with three half-bound staples and another four states with one fully bound and two half-bound staples. Finally there is the possibility that four half-bound staples are attached to the template. For a single-domain staple and the associated pair of template domains there are just four states. There are therefore $34^x \times 4^y$ states of the dimer template with

staples, including part-folded states, where x is the number of two-domain staples and y is the number of single-domain staples. Of these, 2^x states consist exclusively of fully-bound staples. Formally, the state space S is given by $p_0 \times p_1 \times \dots \times p_{k-1}$ where p_i denotes the set of possible states for staple i as described above and k is the total number of staples.

Exclusion algorithm. Two template domains hybridized to a single two-domain staple are held within a few tenths of a nanometre of each other at the staple crossover: many of the folds in S cannot meet this constraint. We provide an algorithm that provides an approximate representation of steric constraints, preventing the model from accessing unrealistic states. This method provides an approximation to the real steric constraints: it does not guarantee that each legal state satisfies the constraints or that all states that satisfy the steric constraints are legal.

We define a connected segment of an origami as a set of hybridized domains such that each domain can be reached from each other domain without leaving the set. Two template domains hybridized to the same staple are defined to be connected, as are two adjacent template domains hybridized to different staples. A partially folded segment of origami is considered stress-free (is legal) when it occurs in one of the set of well-ordered, two-dimensional folds shown in Extended Data Figs 1 or 2. These pre-defined folds satisfy the constraints imposed by finite staple length and steric exclusion.

More formally, we can represent the physical origami in partially folded state $s \in S$ as an abstract graph $G(s) = (V, E)$ such that each boundary between adjacent domains is a vertex $v \in V$ and each template domain and staple crossover is an edge $e \in E$ between the appropriate vertices. Each edge has a labelling function $f: E \rightarrow \{\text{single-stranded, double-stranded, crossover}\}$ that assigns an appropriate status. We can draw subgraphs consisting of connected hybridized segments of the graph: for the origami to be in a legal (stress free) state, each of these subgraphs must be present in a single well-ordered fold from the set shown in Extended Data Figs 1 and 2.

Misfolds occur in the model when at least two connected segments would be incapable of satisfying the constraints were they to become connected to each other. At that point, folding cannot advance unless one of the segments unfolds, allowing another to expand. Extended Data Fig. 3c shows a misfolded dimer that has three connected parts that cannot be joined to form a stress-free state. When simulating assembly using the staple set corresponding to Fig. 4a, about half of the simulations end in a misfolded state; for the weakened-seam variant (Fig. 4b) there are only ~1% misfolds.

Rates model. We develop a kinetic model of folding based on standard reaction models for hybridization and a method to estimate the effective local concentration of the unhybridized domain of a half-bound staple at its complementary template domain.

Consider complementary strands A and B that can bind reversibly to form duplex AB. Under the assumptions of mass action kinetics, the concentration $[AB]$ is described by

$$\frac{d[AB]}{dt} = k_+ [A][B] - k_- [AB] \quad (1)$$

for rate constants k_+ and k_- . The rate constants are constrained by the requirement that the equilibrium concentrations $\{A\}$, $\{B\}$ and $\{AB\}$ are consistent with $\Delta G_{AB}^{0, \text{duplex}}(T)$, the standard change in Gibbs free energy on duplex formation:

$$\frac{\{AB\}}{\{A\}\{B\}} = k_+ / k_- = \exp\left(\frac{-\Delta G_{AB}^{0, \text{duplex}}}{RT}\right) (1M)^{-1} \quad (2)$$

where R denotes the molar gas constant, T temperature.

For staples within a partially folded origami, binding and unbinding rates are similarly constrained by the difference in free energy between states. We approximate the difference in free energy between partially folded states s, s' that differ by the hybridization of a single template domain as

$$\Delta G_{s, s'}^0 = \Delta G_{s, s'}^{0, \text{duplex}} + \Delta G_{s, s'}^{\text{shape}} \quad (3)$$

where $\Delta G_{s, s'}^{0, \text{duplex}}$ is the standard free energy change corresponding to the formation or dissociation of an equivalent isolated duplex and $\Delta G_{s, s'}^{\text{shape}}$ represents the change in entropy corresponding to the geometric constraints on the template that arise when two-domain staples connect non-contiguous template domains ('looping constraints')^{6,30,31}. ΔG^{shape} quantifies cooperative effects: when a single staple domain binds or unbinds, ΔG^{shape} depends on the pattern of binding of other staples.

Consider a single, isolated origami in partially folded state s_{00} and let staple p bind to the template by a single domain, resulting in state s_{01} . The rate for

this reaction is taken to be equal to that for duplex formation between isolated strands:

$$\sigma(s_{00}, s_{01}) = k_+ [p] \quad (4)$$

where $\sigma(s, s')$ is the rate of transition from state s to s' . The unbinding rate is then determined by a thermodynamic constraint analogous to equation (2):

$$\sigma(s_{01}, s_{00}) = k_+ \exp\left(\frac{\Delta G_{s_{00}, s_{01}}^0}{RT}\right) (1M) = k_+ \exp\left(\frac{\Delta G_{s_{00}, s_{01}}^{0, \text{duplex}}}{RT}\right) (1M) \quad (5)$$

We have set $\Delta G_{s_{00}, s_{01}}^{\text{shape}} = 0$ because transitions $s_{01} \leftrightarrow s_{00}$ do not create or destroy loops in the template. (We do not take into account other ways in which hybridization of a single staple domain affects the free energy of the partly-folded origami, for example, by changing the mechanical properties and thus the free-energy cost of any pre-existing loop of which it forms part.) For the second domain of the staple, once the first domain is bound, we again fix the unbinding rate to be that of the corresponding isolated duplex. This rate does not depend on the change in entropy that results from the removal of a looping constraint^{30,31} because, immediately after unbinding, the conformation of the template is unchanged:

$$\sigma(s_{11}, s_{01}) = k_+ \exp\left(\frac{\Delta G_{s_{01}, s_{11}}^{0, \text{duplex}}}{RT}\right) (1M) \quad (6)$$

where s_{11} denotes the state in which the staple is bound to the template with both domains. The binding rates of the second domain of the staple, once the first domain is bound, can then be found from the thermodynamic constraint

$$\begin{aligned} \sigma(s_{01}, s_{11}) &= \sigma(s_{11}, s_{01}) \exp\left(\frac{-\Delta G_{s_{01}, s_{11}}}{RT}\right) \\ &= k_+ \exp\left(\frac{\Delta G_{s_{01}, s_{11}}^{0, \text{duplex}} - \Delta G_{s_{01}, s_{11}}}{RT}\right) (1M) = k_+ \exp\left(\frac{-\Delta G_{s_{01}, s_{11}}^{\text{shape}}}{RT}\right) (1M) \end{aligned} \quad (7)$$

The free energy penalty $\Delta G_{s, s'}^{\text{shape}}$, that corresponds to the additional geometric constraints associated with the binding of the second staple domain, thus determines the binding rate for the second domain.

Looping constraints. We approximate $\Delta G_{s_{01}, s_{11}}^{\text{shape}} = \Delta G_{s_{01}, s_{11}}^{\text{loop}}$, where $\Delta G_{s_{01}, s_{11}}^{\text{loop}}$ corresponds to the entropic penalty of closing the new loop that forms in the template when the second domain of a staple binds. For other transitions, no loop forms and we take $\Delta G_{s, s'}^{\text{shape}} = 0$. $\Delta G_{s, s'}^{\text{loop}}$ quantifies the difference between the entropic penalties for pinning the template into a loop so that the second staple and template domains can bind and for bringing together two domains unconnected by a loop in a hypothetical ideal system at standard conditions (1 M concentration)³². $\Delta G_{s, s'}^{\text{loop}}$ is thus related to the ratio between the probabilities of bringing two domains into contact in the looped system and in the ideal unconnected system:

$$\Delta G_{s, s'}^{\text{loop}} = RT \ln\left(\frac{P_{\text{loop}}^0}{P_{\text{loop}}^0}\right) \quad (8)$$

Here, P_{loop}^0 is the probability that the origami adopts a confirmation in which the unbound staple arm and the template domain are spontaneously within an interaction radius r_0 of each other, where r_0 is an unspecified small distance necessary for closure of the loop. P_{loop}^0 is the probability that two unconnected molecules would be within r_0 in a hypothetical ideal system of $v^0 = 1/N_A$ litres, N_A being Avogadro's number. The rate of hybridization of a second staple domain is therefore given by

$$\sigma(s_{01}, s_{11}) = k_+ \left(\frac{P_{\text{loop}}^0}{P_{\text{loop}}^0}\right) (1M) = k_+ \left(\frac{P_{\text{loop}}^0}{\frac{4}{3}\pi r_0^3 N_A}\right) \quad (9)$$

so $\left(\frac{P_{\text{loop}}^0}{\frac{4}{3}\pi r_0^3 N_A}\right) (1M)$ denotes the effective concentration of the opposing domain.

As a first approximation we treat the loop of DNA as a freely-jointed chain comprising two types of link, double-stranded DNA and single-stranded DNA (dsDNA and ssDNA respectively). Let $P(r)$ be the probability density for the end-to-end extension of the chain r . Then $P_{\text{loop}}^0 = \int_0^{r_0} P(r) dr$ is the probability that the two domains are separated by at most r_0 .

The end-to-end distance distribution $P(r)$ of a freely-jointed chain, in the limit of a large number of segments, is

$$P(r) = 4\pi r^2 \left(\frac{3}{2\pi E[r^2]}\right)^{3/2} \exp\left(\frac{-3r^2}{2E[r^2]}\right) \quad (10)$$

where $E[r^2]$ is the mean squared distance between the two ends. The result for a single segment type is a classic result of statistical physics^{33,34}. The following

argument shows that the result also holds for a chain with heterogeneous segments. From the central limit theorem, for a large number of segments we expect a Gaussian distribution over the x , y and z components of r . Equation (10) is the only Gaussian distribution that also satisfies the symmetry conditions $E[x] = E[y] = E[z] = 0$, and $E[xy] = E[xz] = E[yz] = 0$.

The internal association rate is therefore given by:

$$\begin{aligned} \sigma(s_{01}, s_{11}) &= k_+ \frac{\int_0^{r_0} 4\pi r^2 \left(\frac{3}{2\pi E[r^2]}\right)^{3/2} \exp\left(\frac{-3r^2}{2E[r^2]}\right) dr}{\frac{4}{3}\pi r_0^3 N_A} \\ &\approx k_+ \frac{\int_0^{r_0} 4\pi r^2 \left(\frac{3}{2\pi E[r^2]}\right)^{3/2} dr}{\frac{4}{3}\pi r_0^3 N_A} \approx \frac{k_+}{N_A} \left(\frac{3}{2\pi E[r^2]}\right)^{3/2} \end{aligned} \quad (11)$$

where we have assumed $r_0 \ll E[r^2]$ in the second step.

The loop that is closed by the insertion of a staple into a partly-folded origami has, in general, a complex structure comprising multiply connected domains of single- and double-stranded DNA. We approximate this loop by a single path through the origami, the loop with the smallest expected square end-to-end distance $E[r^2]$. This path represents the most important constraint that leads to the enhancement of the effective local concentration of one end of the loop at the other, and thus provides the most significant enhancement of $\sigma(s_{01}, s_{11})$. In order to identify the dominant loop, each edge $e \in E$ in the implied graph $G(s) = (V, E)$ of the partially folded origami is assigned a weight equal to the contribution to $E[r^2]$ in the freely jointed chain approximation. Dijkstra's shortest path algorithm³⁵ is used to determine a loop that minimizes $E[r^2]$ and hence determines $\sigma(s_{01}, s_{11})$.

For the seam staples, which are paired, the loop closed by hybridization of the second staple is particularly small: it consists only of the crossover link. The predictions of the model remain physically sensible: a second staple binding to a seam has an overall ΔG which is ~ 4.4 kcal mol⁻¹ less favourable (at $T = 60^\circ\text{C}$) than a continuous duplex. This destabilization is equal to that expected from a 5-nt bulge within a duplex³⁰. We note that for the broken seam variant, the model predicts incorporation temperatures for the unbroken staple that are lower than the regular case by 2.0°C , compared to 2.2°C measured in experiment (Extended Data Fig. 6). It is therefore clear that we do not overestimate the cooperative stabilization of seam staples.

The approximations made in estimating the change in free energy when a staple domain binds or unbinds are not thermodynamically self-consistent: the value assigned to the difference in free energy between states depends, in general, on the path taken between them. Models of this kind will be presented in a companion paper, in which they are compared to thermodynamically self-consistent approaches for simpler systems (F.D. *et al.*, submitted).

Parameterization of the model. Compared to unbinding rates, the rate of binding of an isolated duplex is known to be weakly dependent on duplex stability³⁶. We assume k_+ to be independent of temperature, domain sequence, and folding state, and we set $k_+ = 10^6$ M⁻¹ s⁻¹ (refs 25, 36, 37).

The free energy change when each domain binds to its complement, $\Delta G_{s, s'}^{0, \text{duplex}}$, is taken to be that of a 16-bp DNA duplex averaged over all possible sequences³⁸. Buffer conditions of 40 mM [Tris] and 12.5 mM [Mg²⁺] are assumed, giving an additional entropic penalty (in units of cal mol⁻¹ K⁻¹) for duplex formation of:³⁸⁻⁴⁰

$$\Delta S^{0, \text{salt}} = 0.368 \times \frac{N}{2} \times \ln\left(\frac{1}{2} [\text{Tris}] + 3.3 [\text{Mg}^{2+}]^{1/2}\right) \quad (12)$$

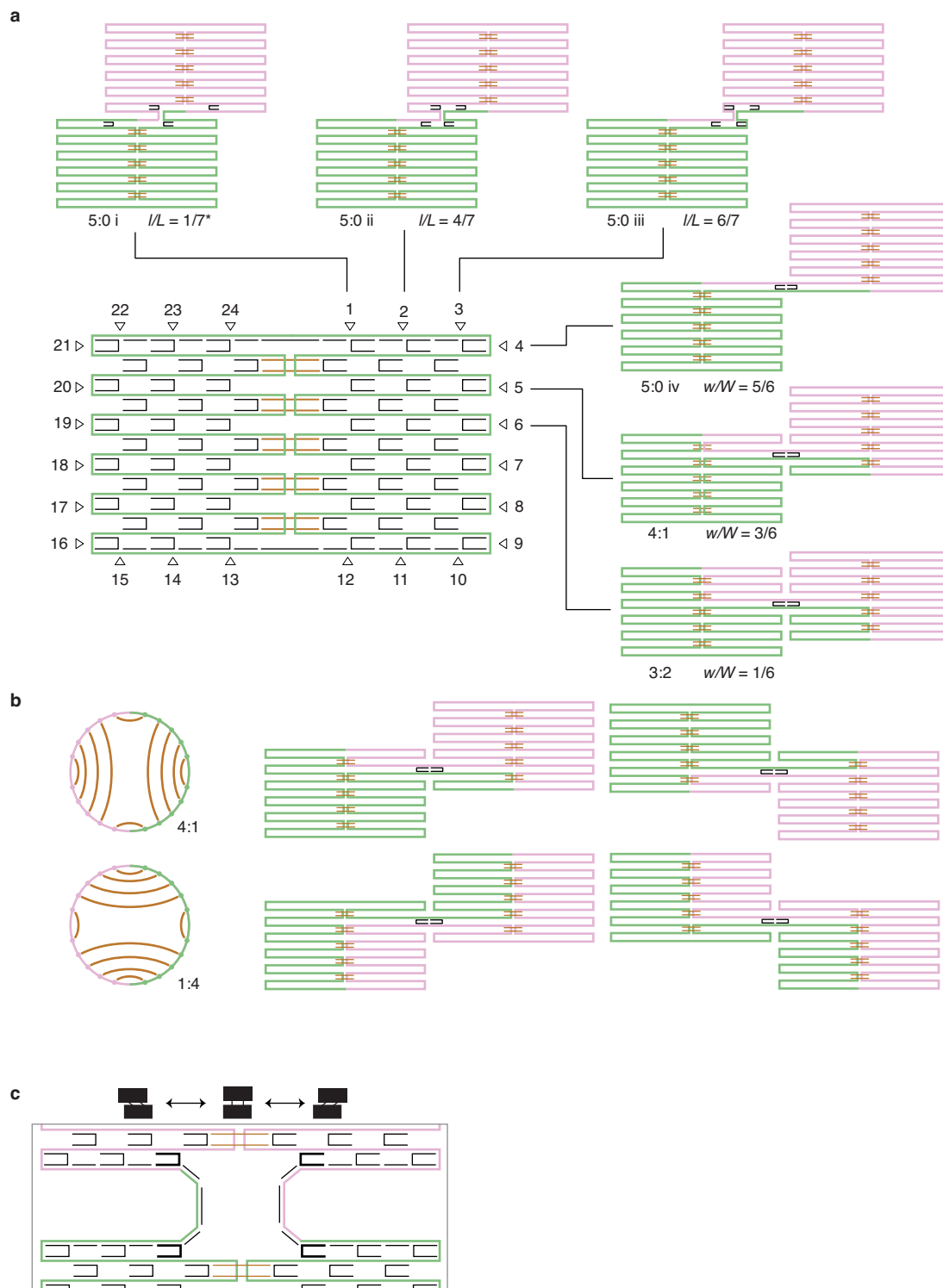
where N is the number of phosphates in the duplex. For ssDNA we use a contour length of $L_{\text{css}} = 0.6$ nm per base and a Kuhn length of $\lambda_{\text{ss}} = 1.8$ nm: a single-stranded domain of 16 bases thus has a contour length of 16×0.6 nm⁴¹⁻⁴⁵. For dsDNA we use a contour length of $L_{\text{c,ds}} = 0.34$ nm per base⁴⁶ and make the approximation that the persistence length is much longer⁴⁷ than any relevant duplex: a double-stranded domain of 16 bases thus corresponds to a single rigid link of length $\lambda_{\text{ds}} = 16 \times 0.34$ nm. A crossover link between the two template domains hybridized to a single staple is treated as a single segment of length λ_{ss} .

Example rate calculations. Consider the half-bound staple shown in Extended Data Fig. 10a that is hybridized to an otherwise empty template. A seam staple, labelled A, is used as an example here. Its second domain can hybridize to either of two sites: the closer is connected by a 448-nt ssDNA chain ($E[r^2] = 480$ nm²) and the further by a composite chain comprising a 2,208-nt single-stranded chain and one rigid 16-bp double stranded segment ($E[r^2] = 2,400$ nm²). Following the calculation outlined above, we find that for the closer site the effective local concentration of the opposing domain $c_{\text{eff}} = 51$ μM , the loop cost $\Delta G_{s, s'}^{\text{loop}} = 6.5$ kcal mol⁻¹ (at $T = 60^\circ\text{C}$) and the hybridization rate $\sigma = 50$ s⁻¹. For the further site: $c_{\text{eff}} = 4.6$ μM , $\Delta G_{s, s'}^{\text{loop}} = 8.1$ kcal mol⁻¹, and $\sigma = 4.5$ s⁻¹. The staple is 11 times more likely to bind to the closer domain.

Binding of one staple affects the binding of others by changing the characteristics of the template (or partly-formed origami) that links their two binding domains. We now compute the hybridization rate, loop cost and local concentration for a second seam staple, staple B, in the presence or absence of staple A. In the absence of staple A, the shorter of the two loops that connect two binding domains of the second staple consists of a 864-nt ssDNA chain: $E[r^2] = 980 \text{ nm}^2$, $c_{\text{eff}} = 18 \mu\text{M}$, $\Delta G^{\text{loop}} = 7.2 \text{ kcal mol}^{-1}$, $\sigma = 18 \text{ s}^{-1}$. In the presence of staple A, the loop passes through the link formed by staple A and comprises 384 nt ssDNA, 3 rigid 16-bp dsDNA segments and a staple crossover modelled as a single segment of length λ_{ss} (Extended Data Fig. 10b): for this shortened loop, $E[r^2] = 520 \text{ nm}^2$, $c_{\text{eff}} = 46 \mu\text{M}$, $\Delta G^{\text{loop}} = 6.6 \text{ kcal mol}^{-1}$ and $\sigma = 46 \text{ s}^{-1}$. Insertion of staple A increases the rate of hybridization of the second domain of staple B by a factor of 2.6 by shortening the distance between its binding sites.

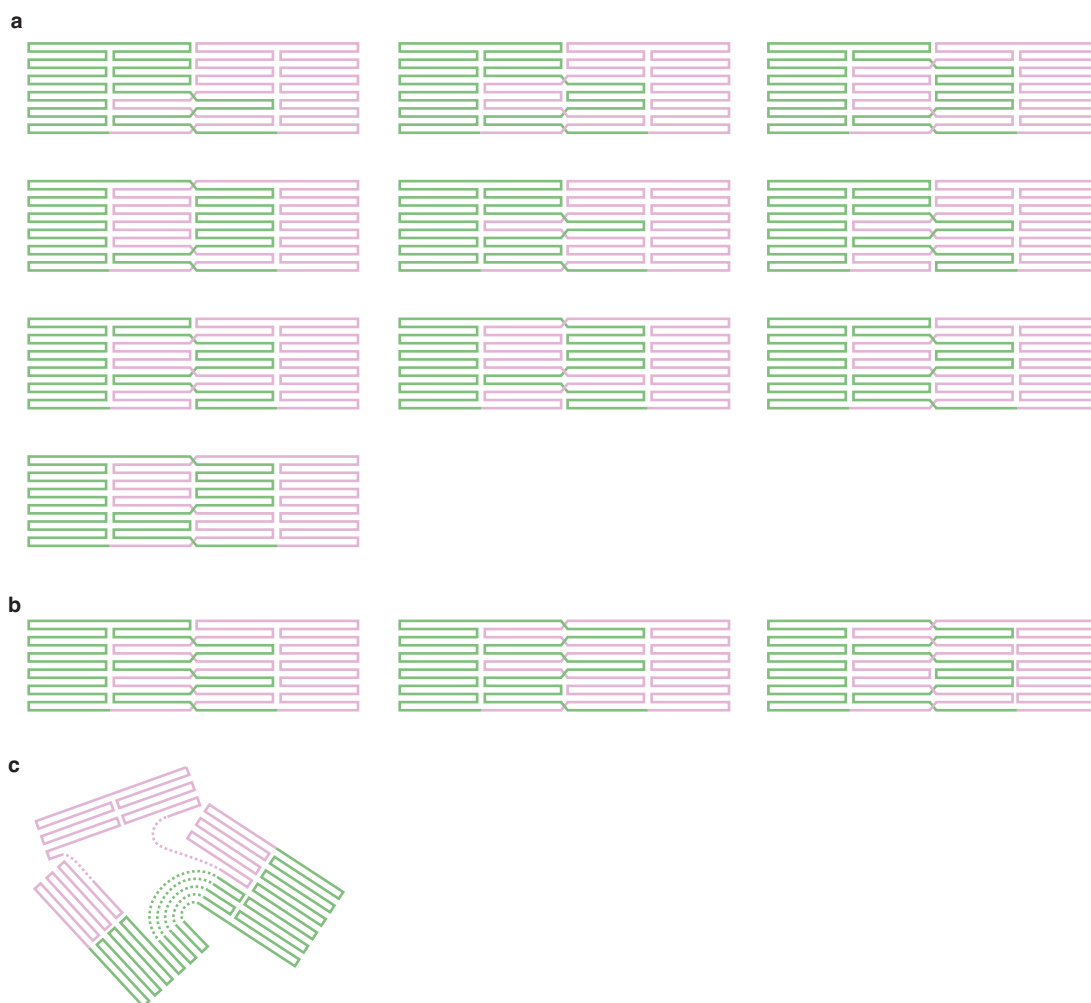
Code availability. The code used to implement the folding model is freely available via <https://github.com/fdannenberg/dna>.

26. Zhang, P. H. *et al.* Engineering BspQI nicking enzymes and application of N.BspQI in DNA labeling and production of single-strand DNA. *Protein Expr. Purif.* **69**, 226–234 (2010).
27. Douglas, S. M. *et al.* Rapid prototyping of 3D DNA-origami shapes with caDNAo. *Nucleic Acids Res.* **37**, 5001–5006 (2009).
28. Wickham, S. F. J. *et al.* Direct observation of stepwise movement of a synthetic molecular transporter. *Nature Nanotechnol.* **6**, 166–169 (2011).
29. Gillespie, D. T. Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.* **81**, 2340–2361 (1977).
30. SantaLucia, J. Jr & Hicks, D. The thermodynamics of DNA structural motifs. *Annu. Rev. Biophys.* **33**, 415–440 (2004).
31. Jacobson, H. & Stockmayer, W. H. Intramolecular reaction in polycondensations. I. The theory of linear systems. *J. Chem. Phys.* **18**, 1600–1606 (1950).
32. Ouldridge, T. E., Louis, A. A. & Doye, J. P. K. Extracting bulk properties of self-assembling systems from small simulations. *J. Phys. Condens. Matter* **22**, 104102 (2010).
33. Rayleigh. On the problem of random vibrations, and of random flights in one, two or three dimensions. *Phil. Mag.* **37**, 321–347 (1919).
34. Chandrasekhar, S. Stochastic problems in physics and astronomy. *Rev. Mod. Phys.* **15**, 1–89 (1943).
35. Dijkstra, E. A note on two problems in connexion with graphs. *Numer. Math.* **1**, 269–271 (1959).
36. Morrison, L. E. & Stols, L. M. Sensitive fluorescence-based thermodynamic and kinetic measurements of DNA hybridization in solution. *Biochemistry* **32**, 3095–3104 (1993).
37. Gao, Y., Wolf, L. K. & Georgiadis, R. M. Secondary structure effects on DNA hybridization kinetics: a solution versus surface comparison. *Nucleic Acids Res.* **34**, 3370–3377 (2006).
38. SantaLucia, J. Jr. A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc. Natl Acad. Sci. USA* **95**, 1460–1465 (1998).
39. Peryet, N. *Prediction of Nucleic Acid Hybridisation: Parameters and Algorithms*. PhD thesis, Wayne State Univ. (2000).
40. Owczarzy, R., Moreira, B. G., You, Y., Behlke, M. A. & Walder, J. A Predicting stability of DNA duplexes in solutions containing magnesium and monovalent cations. *Biochemistry* **47**, 5336–5353 (2008).
41. Smith, S. B., Cui, Y. & Bustamante, C. Overstretching B-DNA: the elastic response of individual double-stranded and single-stranded DNA molecules. *Science* **271**, 795–799 (1996).
42. Rivetti, C., Walker, C. & Bustamante, C. Polymer chain statistics and conformational analysis of DNA molecules with bends or sections of different flexibility. *J. Mol. Biol.* **280**, 41–59 (1998).
43. Mills, J. B., Vacano, E. & Hagerman, P. J. Flexibility of single-stranded DNA: use of gapped duplex helices to determine the persistence lengths of poly(dT) and poly(dA). *J. Mol. Biol.* **285**, 245–257 (1999).
44. Murphy, M. C., Rasnik, I., Chang, W., Lohman, T. M. & Ha, T. Probing single-stranded DNA conformational flexibility using fluorescence spectroscopy. *Biophys. J.* **86**, 2530–2537 (2004).
45. Chen, H. *et al.* Ionic strength-dependent persistence lengths of single-stranded RNA and DNA. *Proc. Natl Acad. Sci. USA* **109**, 799–804 (2012).
46. Saenger, W. *Principles of Nucleic Acid Structure* (Springer, 1984).
47. Hagerman, P. J. Flexibility of DNA. *Annu. Rev. Biophys. Biophys. Chem.* **17**, 265–286 (1988).



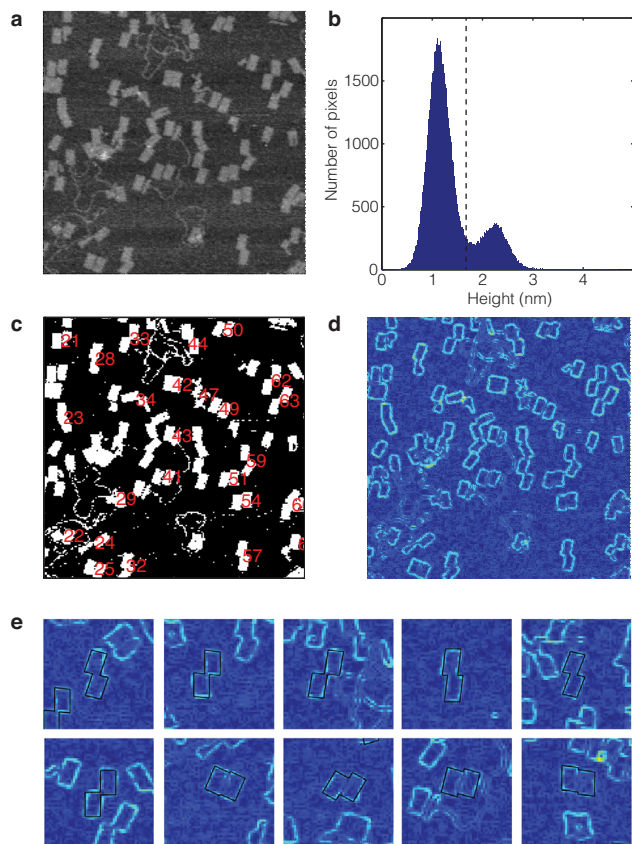
Extended Data Figure 1 | The set of well-folded, planar states. **a**, Well-folded, planar states can be considered as two adjacent monomer tiles linked by a single reciprocal template crossing at any of the locations marked with a triangle and numbered (centre). This gives a set of 6 unique shapes, as indicated (periphery). **b**, With the exception noted below, there are four ways to make each of these shapes, distinguished by the nucleotide sequence at the template crossing but not resolved by AFM imaging. In the example shown, crossings made at positions 5 and 8 correspond to the fold 4:1, and crossings made at 17 and 20 correspond to fold 1:4, as indicated in the circle diagrams (left). All give the same shape with fractional short edge offset w/W of $3/6$ (right).

The exception is that there are only two variants of state 5:0i as configurations formed by linking tiles at positions 1 and 24 are not distinguishable, nor are links at positions 12 and 13. **c**, Detailed view of the connection between monomer tiles in this case, for which the long-edge offset is not precisely defined (it can range from 0 to $2/7$ depending on the conformation of the long edge staple). For the purpose of predicting geometry for model configurations, we take an average value of $l/L = 1/7$. The set of 22 well-folded, planar states thus consists of two folds for the shape shown in **c** and four folds for each of the other five shapes.

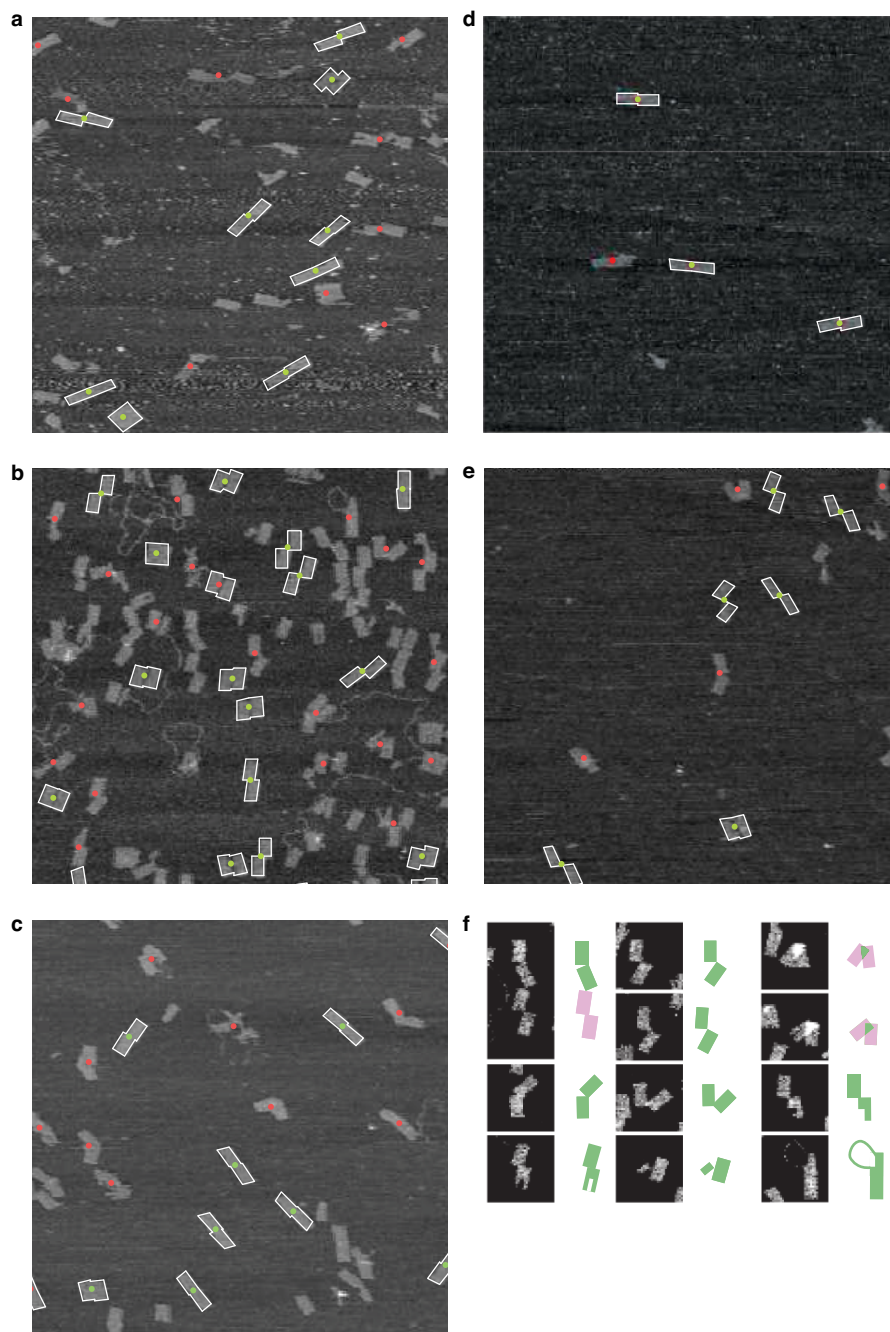


Extended Data Figure 2 | Well-folded, non-planar states and an illegal fold.
a, b, The set of legal folds permitted by the model consists of the 22 planar folds defined in Extended Data Fig. 1 and an additional 52 non-planar folds, four for each of the 13 shapes shown here in **a, b**. Shapes in **a** are formed by

allowing three reciprocal crossings between two tiles, those in **b** are formed by allowing 5 reciprocal crossings. These non-planar folds form only rarely in simulation. **c,** An example of a misfolded shape: the part-folded domains are, individually, well-formed but cannot be joined to give a legal fold.

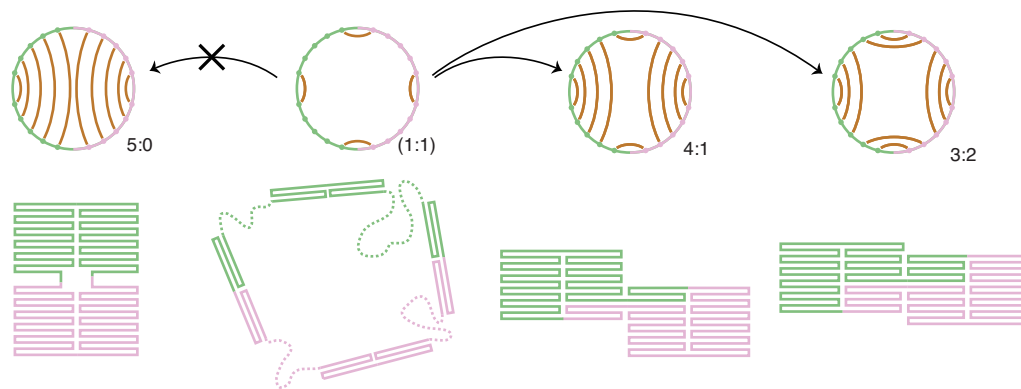


Extended Data Figure 3 | Fitting the shapes of origami tiles observed by AFM. **a**, AFM images were flattened by line-by-line subtraction of a second-order polynomial. Image processing and fitting were performed using custom MATLAB programs. Image $1.2 \times 1.2 \mu\text{m}$. **b**, A histogram of pixel heights was used to set the threshold for the generation of a binary image. The threshold was found by calculating the average of the means of the two peaks corresponding to background and tiles; if this failed because the image was noisy the threshold was set manually. **c**, Well-separated objects in the binary image which have the approximate area of a dimer tile were flagged for fitting (numbered). **d**, Tile outlines were generated using a Sobel edge-finding filter. **e**, Representative fitted outlines (two equal, offset parallelograms) were used to classify dimer tiles as described in the text (compare Fig. 3d).



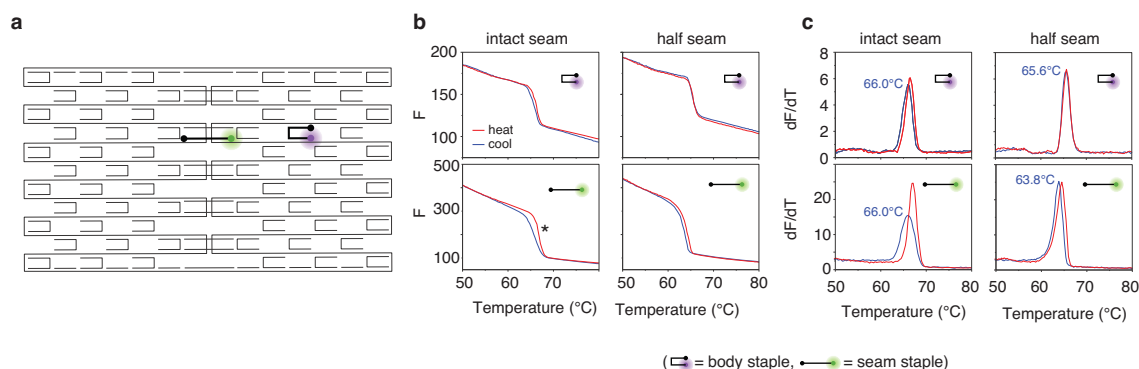
Extended Data Figure 4 | AFM data. Panels **a–e** show a 1.5 μm field of view containing structures folded from each of the five staple sets of Fig. 4a–e. Shapes that were flagged for fitting are marked with a dot, green if the shape was successfully fitted and red otherwise. The fitted outlines are superimposed on the image. **f**, Examples of structures that were either not flagged for fitting or not successfully fitted. AFM images are shown alongside the outline of a suggested structure. The collection of shapes that were not successfully fitted

includes crowded areas where shapes are touching and shapes where the two component monomer tiles are distorted, perhaps during deposition on the mica surface, but can be clearly assigned to one of the predicted shapes. Part-folded (or damaged) shapes are also observed, often with one well-folded monomer attached to a part-folded monomer; sometimes a portion of unfolded template can be observed.



Extended Data Figure 5 | Strong seam connections influence the folding pathway. The structure labelled 1:1 is a part-folded intermediate in which four pairs of seam staples are bound. If the seam staples remain in place this intermediate could progress to a fully folded structure with seam configuration

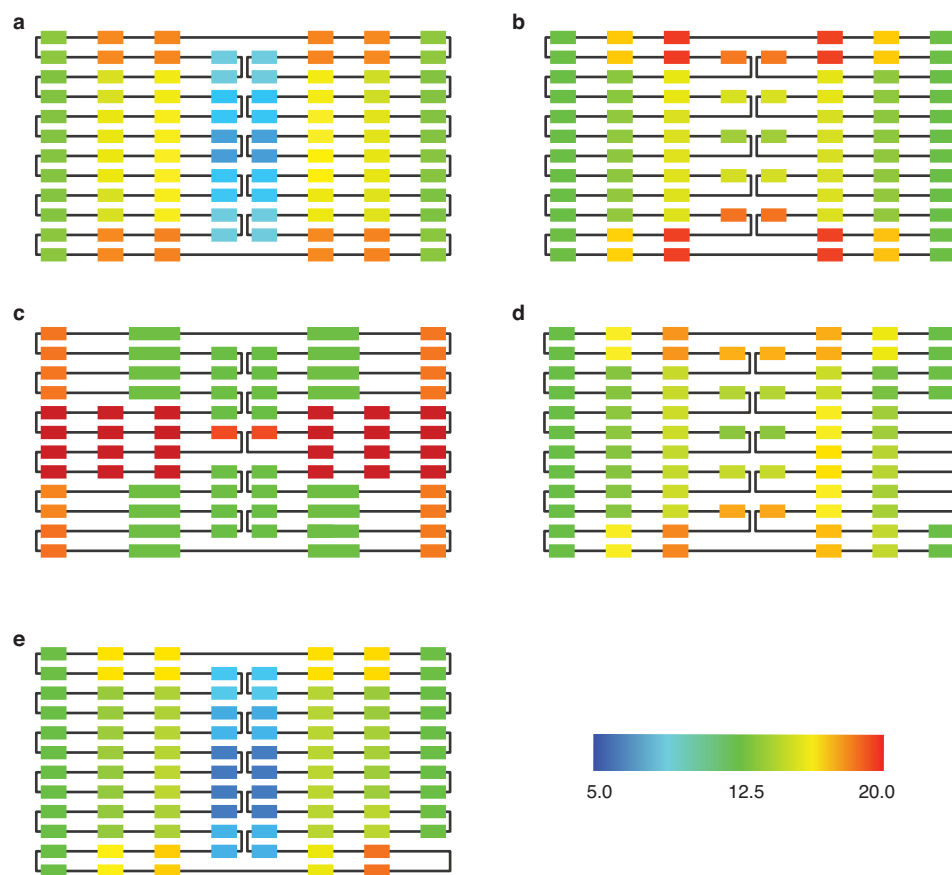
4:1 or 3:2 (as indicated by arrows to the right), but fold 5:0 is inaccessible unless two pairs of seam staples dissociate. Circle diagrams in the upper panel show seam connections corresponding to the structures below.



Extended Data Figure 6 | Monitoring origami assembly using fluorescence.

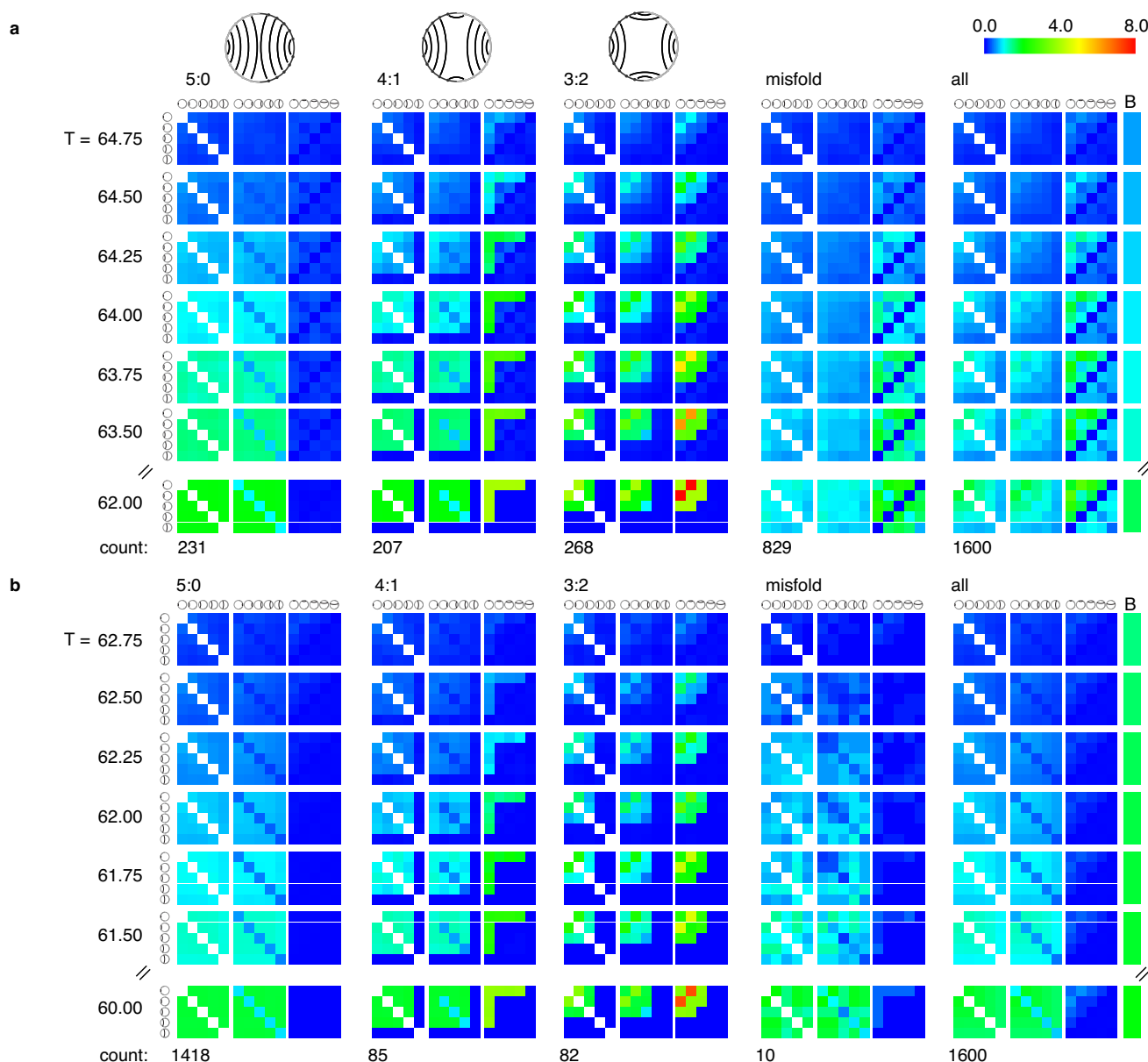
Assembly of a monomer tile (Fig. 1) was monitored using fluorescently labelled staples. The positions of the labelled strands in the folded tile are shown in **a**: the seam staple was labelled with 5' Cy3 and 3' Black Hole Quencher 2, and the body staple with 5' Cy5 and 3' Black Hole Quencher 2. Reactions containing the monomer template at 50 nM and staples at 100 nM in a buffer containing 12.5 mM MgCl₂, 10 mM Tris-HCl and 0.5 mM EDTA pH 8.0 were held at 96 °C for 10 min, cooled from 96 °C to 25 °C at 0.3 °C min⁻¹, held at 25 °C for 10 min then heated to 96 °C at 0.3 °C min⁻¹. The fluorescence signal for Cy3 and Cy5 was recorded at 0.3 °C intervals during cooling and heating cycles. Staple binding increases the separation between fluorophore and

quencher and therefore increases the fluorescence intensity. **b**, Fluorescence intensities (*F*) and **c**, their derivatives (*dF/dT*) as functions of temperature during origami annealing and melting. Sharp transitions, corresponding to narrow ranges of staple incorporation temperatures, are consistent with cooperative origami assembly. In the case of the unmodified tile the seam staple is incorporated into the tile at the same temperature as the body staple. Hysteresis (marked *) is consistent with the cooperative binding of the seam staple. When one half of the seam is broken the hysteresis observed for seam staple binding is reduced and the seam staple is incorporated at a lower temperature than the body staple. Weakening the seam has little effect on the incorporation of the body staple.



Extended Data Figure 7 | Rearrangement of staples during folding.
a–e, Heat maps showing the predictions of the model for the number of reconfiguration events during assembly for each of the staple sets shown in Fig. 4a–e. A ‘reconfiguration event’ occurs when a contact between two







template domains is released and replaced by an alternative contact. Domains omitted from the map are those which would generate an illegal fold if reconfigured.



Extended Data Figure 8 | Evolving correlations between seam staples in the model during folding.

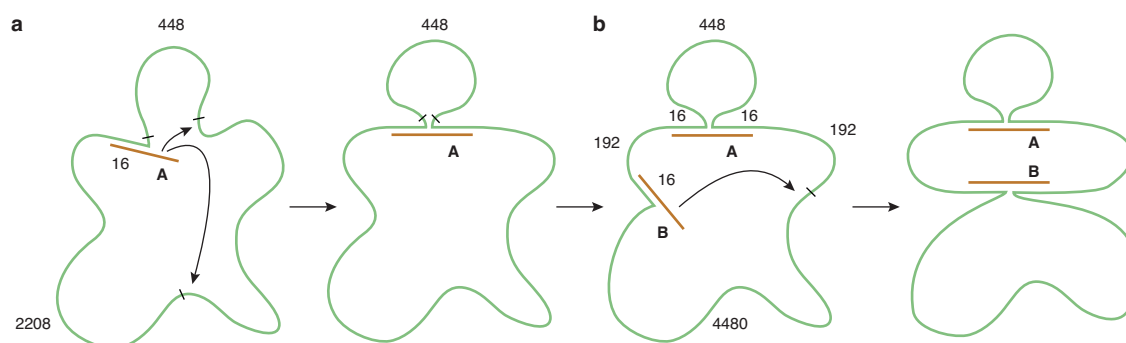
a, The original staple set (Fig. 4a); **b**, the broken-seam variant (Fig. 4b). In each case, average data from 1,600 simulations are presented ('all') together with subsets sorted by final fold (5:0, 4:1, 3:2 and misfold). The simulation count for each subset is indicated below each panel. Simulations resulting in well-folded, non-planar structures (NP) are included in 'all' but not presented separately: such structures occurred 65 times in **a** and 5 times in **b**. Circular icons with internal connections of different lengths represent links across the seam ('seam links') connecting points on the template spanning (that is, that are separated by) 28, 56, 84, 112 and 140 template domains (as in the 'circle' diagrams of Fig. 3). A 'seam link' represents a connection across the seam mediated by at least one seam staple (with the original staple set, **a**, it may also represent a pair of staples). Data are presented at seven different temperatures as the system is cooled. Correlations between seam links are represented graphically by three 5×5 blocks. Each pixel represents a correlation between a pair of seam links which are identified by two icons. The orientation of the icons has the same significance as in the 'circle' diagrams: two icons related by 180° rotation represent one internal link in each of the two halves of the template; a 90° rotation represents one internal link and one cross link. Only relative orientation is significant so, for example, fully folded state $m:n$ is not distinguished from $n:m$ (Extended Data Fig. 1b). Each pixel represents the average number of pairs of links present with the specified spans and relative orientations (range 0–8; colour coded, key at top right). The bar on the right of the figure, labelled 'B', represents the average occupancy of body staples

(range 0–2). For staple set **a**, folding is substantially complete at 62°C : at this temperature the patterns of correlation that are characteristic of the fully-folded structures can be seen clearly. For example, the presence of the longest (140-domain) link with no cross-link to the other half of the template is characteristic of fold 5:0. (A 140-domain link with a cross-link only occurs in misfolds and NP structures.) A 112-domain link with a 28-domain cross-link is characteristic of 4:1, and the presence of two 56-domain links including a cross-link is characteristic of 3:2. These and other correlations that are characteristic of the final folds are already visible in the averaged correlation maps (when simulations are sorted by final fold) at very early stages of folding. The pattern of seam staples at an early stage of folding is therefore predictive of the final fold (Extended Data Fig. 9). For the broken-seam staple set **b**, intact seam staples are incorporated later in the folding pathway (the 50% incorporation temperature for seam staples is 64.2°C for **a**, 62.3°C for **b**). The 50% body staple incorporation temperature is unchanged (63.9°C for **a**, 64.0°C for **b**). The same characteristic patterns of seam staples that, with the full seam, are associated with different final folds are also visible at high temperatures for the broken-seam staples. However, 90% of broken-seam simulations result in fold 5:0, as designed. Additional evidence for the influence of strong seam contacts on the folding pathway in the model is provided by the dramatically different yields of misfolds: 52% for full-seam staples **a**, 1% for broken-seam staples **b**. Stable incorporation of incompatible seam staples in **a** prevents the formation of well-folded structures.

Test	5:0	4:1	3:2	N.P.	misfold	all
1.  and not 	117	8	5	5	82	217
2.  and 	16	100	28	12	142	298
3.  and 	20	19	155	11	176	381
Total:	231	207	268	65	829	1600

Extended Data Figure 9 | Seam–staple correlations at early stages of folding are predictive of the final fold. Data shown correspond to the original staple set (see Extended Data Fig. 8a and Fig. 4a). Three tests were applied at the temperature at which, on average, half of all seam staples are incorporated (64.2 °C). These tests were designed to discriminate between patterns of seam staples characteristic of different final folds. For simulations that satisfy each test, the table records the distribution between final folds. Test 1: a 140-domain

seam link with no cross-link to the other half of the template (characteristic of fold 5:0). Test 2: a 112-domain link with a 28-domain cross-link (characteristic of fold 4:1). Test 3: two 56-domain links, including one internal link and one cross-link between halves of the template (characteristic of fold 3:2). Highlighted entries correspond to the fold that each test was designed to predict. The last row of the table records the final distribution between folds of all 1,600 simulations.



Extended Data Figure 10 | Example calculations of staple hybridization rates. See Methods section ‘Example rate calculations’ for the worked examples. **a**, A half-bound seam staple (brown) can bind to one of two sites on the template (green). Distances along the template to each of the two possible binding sites for the second domain of the staple, measured in nucleotides and base pairs, are marked on the template. In the example shown, the closer binding site is connected by a 448-nt ssDNA chain and the further by a composite chain comprising a 2,208-nt single-stranded chain and one rigid 16-bp double stranded segment. The local concentration of the closer domain at the half-bound staple is estimated to be 11 times higher than that of the more

distant domain with a correspondingly greater hybridization rate. **b**, The previous incorporation of staples changes the physical properties of the loops connecting staple binding sites and thus staple incorporation rates. In the absence of staple A, the shortest path between the binding domains of staple B shown consists of a 864-nt ssDNA chain. In the presence of staple A the path is shortened: it passes through the link formed by staple A and comprises 384 nt ssDNA, 3 rigid 16-bp dsDNA segments and a staple crossover. The effect of the previous insertion of staple A, shortening the link between the two binding sites, is to accelerate the hybridization of the second domain of staple B by a factor of 2.6.

Alcohols as alkylating agents in heteroarene C–H functionalization

Jian Jin¹ & David W. C. MacMillan¹

Redox processes and radical intermediates are found in many biochemical processes, including deoxyribonucleotide synthesis and oxidative DNA damage¹. One of the core principles underlying DNA biosynthesis is the radical-mediated elimination of H₂O to deoxygenate ribonucleotides, an example of ‘spin-centre shift’², during which an alcohol C–O bond is cleaved, resulting in a carbon-centred radical intermediate. Although spin-centre shift is a well-understood biochemical process, it is underused by the synthetic organic chemistry community. We wondered whether it would be possible to take advantage of this naturally occurring process to accomplish mild, non-traditional alkylation reactions using alcohols as radical precursors. Because conventional radical-based alkylation methods require the use of stoichiometric oxidants, increased temperatures or peroxides^{3–7}, a mild protocol using simple and abundant alkylating agents would have considerable use in the synthesis of diversely functionalized pharmacophores. Here we describe the development of a dual catalytic alkylation of heteroarenes, using alcohols as mild alkylating reagents. This method represents the first, to our knowledge, broadly applicable use of unactivated alcohols as latent alkylating reagents, achieved via the successful merger of photoredox and hydrogen atom transfer catalysis. The value of this multi-catalytic protocol has been demonstrated through the late-stage functionalization of the medicinal agents, fasudil and milrinone.

During DNA biosynthesis, ribonucleoside diphosphates are converted into their deoxyribonucleoside equivalents via the enzymatic activity of ribonucleotide reductase (class I–III)⁸. Crucially, a (3′,2′)-spin-centre shift occurs, resulting in β-C–O scission and elimination of water (Fig. 1a). Considering the efficiency of this mild enzymatic process to cleave C–O bonds to generate transient radicals, we postulated whether an analogous chemical process could occur with simple alcohols, such as methanol, to access radical intermediates for use in challenging bond constructions (Fig. 1b). In the medicinal chemistry community, there is growing demand for the direct introduction of alkyl groups, especially methyl groups, to heteroarenes, given their influence on drug metabolism and pharmacokinetic profiles⁹. The open-shell addition of alkyl radical intermediates to heteroarenes, known as the Minisci reaction¹⁰, has become a mainstay transformation with broad application within modern drug discovery¹¹. Unfortunately, many current methods are limited in their application to late-stage functionalization of complex molecules owing to their dependence on the use of strong stoichiometric oxidants or increased temperatures to generate the requisite alkyl radicals^{3–6}. A photoredox-catalysed alkylation protocol using peroxides as the alkyl radical precursors was recently demonstrated⁷. Given the state of the art, we questioned whether a general alkylation protocol could be devised in which a broad range of substituents could be installed from simple commercial alcohols under mild conditions.

Visible light-mediated photoredox catalysis has emerged in recent years as a powerful technique in organic synthesis that facilitates single-electron transfer events with organic substrates^{12–14}. This general strategy allows for the development of bond constructions that are often elusive or currently impossible via classical two-electron

pathways. Recently, our laboratory introduced a new dual photoredox-organocatalytic platform to enable the functionalization of unactivated *sp*³ C–H bonds^{15–17}. This catalytic manifold provides access to radical intermediates via C–H abstraction, resulting in the construction of challenging C–C bonds via a radical–radical coupling mechanism. With the insight gained from this dual catalytic system and our recent work on the development of a photoredox-catalysed Minisci reaction¹⁸, we questioned whether it would be possible to generate alkyl radicals from alcohols and use them as alkylating agents in a heteroaromatic C–H functionalization reaction (Fig. 1c). While there are a few early reports of alcohols as alkyl radical precursors formed via high-energy irradiation (ultraviolet light and gamma rays)^{19–21}, a general and robust strategy for using alcohols as latent alkylating agents has been elusive. This transformation would represent a direct C–H alkylation of heteroarenes with alcohols via a spin-centre shift pathway, eliminating H₂O as the only by-product. We recognized that this mild alkylating procedure would serve as a powerful and general method in late-stage functionalization, using commercially available and abundant alcohols as latent alkylating agents.

A detailed description of our proposed dual catalytic mechanism for the alkylation of heteroarenes with alcohols is outlined in Fig. 2. Irradiation of Ir(ppy)₂(dtbbpy)⁺ (**1**) (in which ppy = 2-phenylpyridine, dtbbpy = 4,4′-di-*tert*-butyl-2,2′-bipyridine) will generate the long-

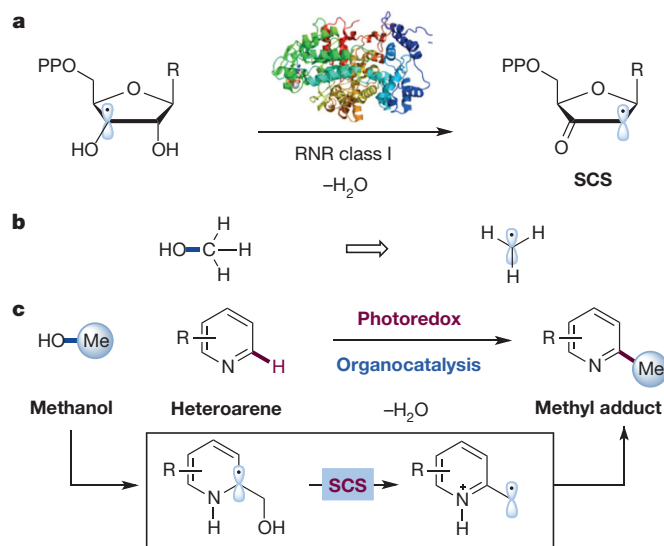


Figure 1 | Bio-inspired alkylation process using alcohols as spin-centre shift equivalents via a dual catalytic platform. **a**, DNA biosynthesis occurs via a spin-centre shift (SCS) process, catalysed by ribonucleotide reductase (RNR) class I to generate a carbon-centred radical, after elimination of H₂O as a by-product. **b**, Alcohols (for example, methanol) as radical intermediates when spin-centre shift allowed. **c**, Proposed direct installation of alkyl groups using alcohols under mild photoredox organocatalytic conditions.

¹Merck Center for Catalysis at Princeton University, Princeton, New Jersey 08544, USA.

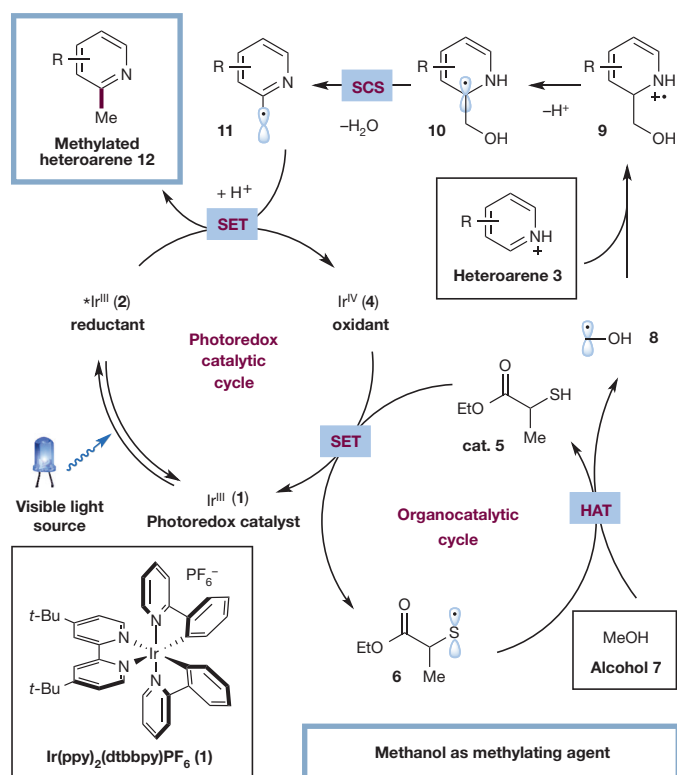


Figure 2 | Proposed mechanism for the direct alkylation of heteroaromatic C-H bonds via photoredox organocatalysis. The catalytic cycle is initiated via excitation of photocatalyst **1** to give the excited state **2**. A sacrificial amount of heteroarene **3** oxidizes $^*\text{Ir}^{\text{III}}$ **2** to Ir^{IV} **4**, which then oxidizes thiol catalyst **5** to generate thiyl radical **6** and regenerate catalyst **1**. Thiyl radical **6** then abstracts a hydrogen atom from alcohol **7** to form α -oxy radical **8**. Radical **8** adds to heteroarene **3**, producing radical cation **9**, which after deprotonation forms α -amino radical **10**. Spin-centre shift elimination of H_2O forms radical intermediate **11**. Protonation and reduction by $^*\text{Ir}^{\text{III}}$ **2** delivers alkylated product **12**. HAT, hydrogen atom transfer; MeOH, methanol; SET, single-electron transfer.

lived $^*\text{Ir}(\text{ppy})_2(\text{dtbbpy})^+$ (**2**) excited state ($\tau = 557 \text{ ns}$)²². As $^*\text{Ir}(\text{ppy})_2(\text{dtbbpy})^+$ (**2**) can function as either a reductant or an oxidant, we postulated that **2** would undergo a single-electron transfer event with a sacrificial quantity of protonated heteroarene **3** to initiate the first catalytic cycle and provide the oxidizing $\text{Ir}(\text{ppy})_2(\text{dtbbpy})^{2+}$ (**4**). Given the established oxidation potential of $\text{Ir}(\text{ppy})_2(\text{dtbbpy})^{2+}$ (**4**) ($E_{1/2}^{\text{red}} = +1.21 \text{ V}$ versus saturated calomel electrode in CH_3CN)²², we anticipated that single-electron transfer from the thiol catalyst **5** ($E_{1/2}^{\text{red}} = +0.85 \text{ V}$ versus saturated calomel electrode for cysteine)²³ to $\text{Ir}(\text{ppy})_2(\text{dtbbpy})^{2+}$ (**4**) would occur and, after deprotonation, furnish the thiyl radical **6** while returning $\text{Ir}(\text{ppy})_2(\text{dtbbpy})^+$ (**1**) to the catalytic cycle. At this stage, we presumed that the thiyl radical **6** would undergo hydrogen atom transfer with the alcohol **7** (a comparable thiol, methyl 2-mercaptoacetate S-H bond dissociation energy = 87 kcal mol^{-1} (ref. 24), methanol α -C-H bond dissociation energy = 96 kcal mol^{-1} (ref. 25)) to provide the α -oxy radical **8** and regenerate the thiol catalyst **5**, driven by the polar effect in the transition state²⁶. The polar effect is a remarkable property that enables considerably endergonic C-H abstractions that would not be possible otherwise²⁷. The nucleophilic α -oxy radical **8** would then add to the protonated electron-deficient heteroarene **3** in a Minisci-type pathway to afford the aminyl radical cation **9**. The resulting α -C-H bond of **9** is sufficiently acidic to undergo deprotonation to form the α -amino radical **10** (ref. 28). At this juncture, intermediate **10** is primed to undergo a spin-centre shift to eliminate H_2O and generate benzylic radical **11**. The resulting open-shell species would then undergo protonation followed by a second

single-electron transfer event with the excited photocatalyst **2** to regenerate the active oxidant $\text{Ir}(\text{ppy})_2(\text{dtbbpy})^{2+}$ (**4**), while providing the desired alkylation product **12**.

We first examined this new alkylation protocol using isoquinoline and methanol as the coupling partners, and evaluated a range of photocatalysts and thiol catalysts. Using $\text{Ir}(\text{ppy})_2(\text{dtbbpy})\text{PF}_6$ (**1**) and ethyl 2-mercaptopropionate (**5**), along with *p*-toluenesulfonic acid and blue light-emitting diodes as the light source, we were able to achieve the desired C-C coupling to provide 1-methylisoquinoline (**15**) with a 92% yield (see Supplementary Information). Notably, we observed none of the desired product in the absence of photocatalyst, thiol catalyst, acid or light, demonstrating the requirement of all components in this dual catalytic protocol. In addition, this method requires only weak visible light and ambient temperature to install methyl substituents using methanol as the alkylating agent.

With the optimal conditions in hand, we sought to evaluate the generality of this dual catalytic alkylation transformation. As highlighted in Fig. 3a, a wide range of heteroaromatics are methylated under the reaction conditions. Isoquinolines with electron-donating or -withdrawing substituents (such as methyl substituents, esters and halides) are functionalized in excellent efficiencies (**15–18**, 85–98% yield). Quinolines perform effectively, including those that contain non-participating functionality (**19–23**, 65–95% yield), in addition to phthalazine and phenanthridine coupling partners (**24** and **25**, 70% and 93% yield). Moreover, a wide range of pyridine derivatives containing diverse functionality (such as esters, amides, arenes, nitriles and trifluoromethyl groups) can be converted into the desired methylation products in high yield (**26–32**, 65–91% yield).

Next, we sought to investigate the nature of the alcohol coupling partner, as demonstrated in Fig. 3b. A broad array of primary alcohols can effectively serve as alkylating agents in this new alkylation reaction. In contrast to the methylation conditions highlighted above, alcohols in Fig. 3b typically use methyl thioglycolate **13** as the C-H abstraction catalyst. Notably, simple aliphatic alcohols such as ethanol and propanol deliver the alkylated isoquinoline product in high yields (**33** and **34**, 95% and 96% yield). Steric bulk proximal to the alcohol functionality is tolerated, as exemplified by the presence of isopropyl, β -tetrahydropyran, β -aryl and β -adamantyl substituents (**35–38**, 87–92% yield). The presence of an electron-withdrawing trifluoromethyl (CF_3) group distal to the alcohol decreases the rate of the reaction; however, using the more electrophilic thiol catalyst, 2,2,2-trifluoroethanethiol (**14**), can promote the transformation more efficiently, possibly owing to the polar effect on the hydrogen atom transfer transition state (**39**, 93% yield)²⁶. We found that diols also participate readily in this alkylation protocol (**40** and **41**, 88% and 81% yield). It should be noted that 1,3-butanediol demonstrates exceptional chemoselectivity and undergoes alkylation exclusively at the primary alcohol site. We speculate that the corresponding α -oxy radical at the secondary alcohol position does not attack the protonated heteroarene owing to its increased steric hindrance. For these alkylating agents with several reactive sites (**41**, **43** and **44**), thiol catalyst **5** is the most effective hydrogen atom transfer catalyst—mechanistic studies are continuing to determine the origin of these differences in catalyst reactivity. Ethers, in the form of differentially substituted tetrahydrofurans, are also competent alkylating agents in this dual catalytic platform (**42–44**, 72–90% yield). In the elimination step, the tetrahydrofuran ring opens to reveal a pendent hydroxyl group. Interestingly, 3-hydroxytetrahydrofuran and tetrahydrofurfuryl alcohol react regioselectively at the ether α -oxy site distal to the alcohol to afford alkylation products with terminal pinacol motifs. We attribute this exclusive regioselectivity to a subtle influence on C-H bond dissociation energy owing to the inductive influence of the oxygen atoms. The application of these substrates represents an effective method to install vicinal diol motifs that would be inaccessible using traditional oxidative alkylation methods. Finally, the utility of this mild alkylation protocol has been demonstrated by the late-stage

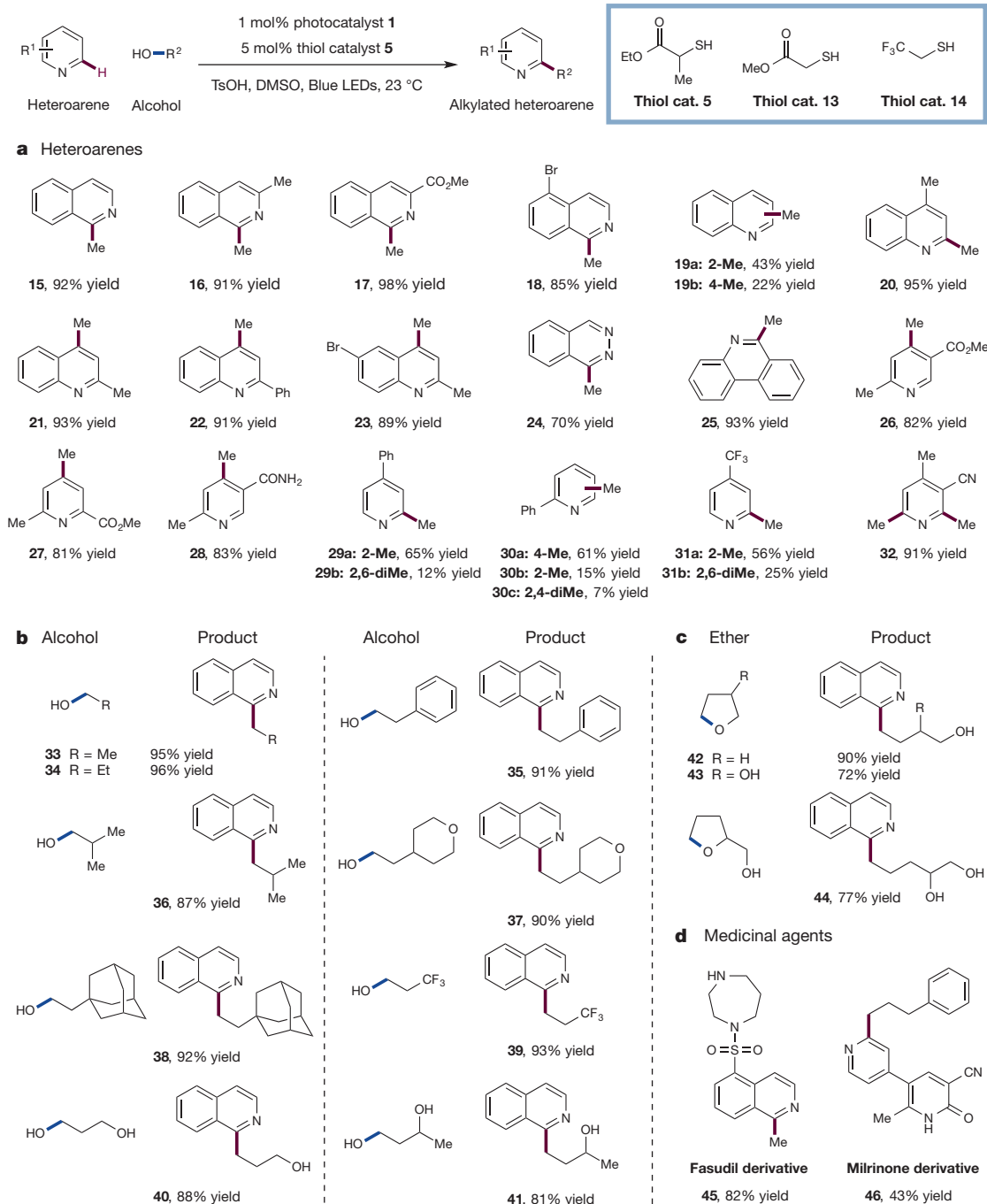


Figure 3 | Substrate scope for the alkylation of heteroaromatic C–H bonds with alcohols via the dual photoredox organocatalytic platform. A broad range of heteroarenes and alcohols are efficiently coupled to produce alkylated heterocycles under the standard reaction conditions (top, generalized reaction). **a**, A variety of isoquinolines, quinolines, phthalazines, phenanthridines and pyridines are efficiently methylated using methanol as the alkylating reagent. **b**, A diverse selection of alcohols serve as effective alkylating

agents in this dual catalytic protocol. **c**, Ethers are also amenable to the transformation; the products are the corresponding ring-opened alcohols. **d**, Two pharmaceuticals, fasudil and milrinone, can be alkylated using this protocol, demonstrating its utility in late-stage functionalization. Isolated yields are indicated below each entry. See Supplementary Information for experimental details.

functionalization of several pharmaceutical compounds. Using methanol as a simple methylating agent, fasudil, a potent Rho-associated protein kinase inhibitor and vasodilator, can be methylated in 82% yield (product **45**). Additionally, milrinone, a phosphodiesterase 3 inhibitor and vasodilator, can be alkylated with 3-phenylpropanol in 43% yield (product **46**).

Mechanistic studies have been conducted to support the proposed pathway outlined in Fig. 2. Stern–Volmer fluorescence quenching experiments have demonstrated that the $^*\text{Ir}^{\text{III}}$ excited state **2** is

quenched in the presence of protonated heteroarene **3**, but not in the presence of the unprotonated heteroarene or thiol catalyst **5**, indicating an oxidative quenching pathway (see Supplementary Information). Furthermore, a series of experiments were conducted to investigate the proposed spin-centre shift elimination. After exposing hydroxylated intermediate **47** to the reaction conditions, only a modest amount of the methylated isoquinoline **15** is observed (8% yield, entry 1, Fig. 4a). In the absence of an acid additive, only trace yields of the desired product are formed (2% yield, entry 2, Fig. 4a). However, in

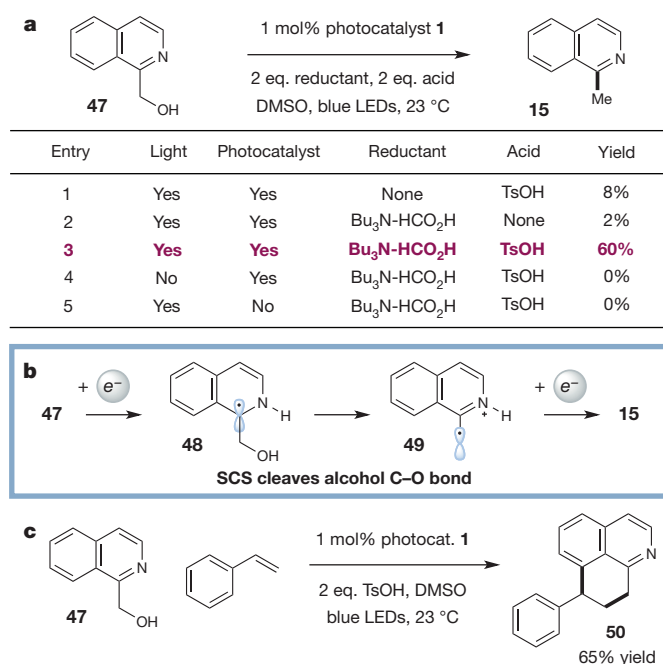


Figure 4 | Mechanistic studies support spin-centre shift elimination pathway. **a**, Hydroxymethyl intermediate **47** can be converted to methylated **15** under net reductive conditions after addition of formic acid-tributylamine and *p*-toluenesulfonic acid (TsOH). **b**, Deoxygenation of **47** probably proceeds via a spin-centre shift pathway to cleave the alcohol C–O bond. **c**, In the presence of styrene, **47** is converted to **50**, presumably by trapping of radical **49**. DMSO, dimethylsulfoxide; LEDs, light-emitting diodes.

the presence of a stoichiometric reductant and *p*-toluenesulfonic acid, the elimination of oxygen can be achieved in good efficiency (60% yield, entry 3, Fig. 4a). Crucially, this elimination pathway is shut down in the absence of either light or photocatalyst (entry 4 or 5, respectively, Fig. 4a). Therefore, this net reductive process supports the proposed generation of α -amino radical **48**, which could readily form deoxygenated product **15** via a spin-centre shift pathway to β -amino radical **49** (Fig. 4b). This elimination pathway is further corroborated by a series of radical trapping experiments (Fig. 4c and Supplementary Information). In the presence of styrene, hydroxymethyl arene **47** is transformed to adduct **50** (65% yield, Fig. 4c), presumably via the intermediacy of β -amino radical **49**. Finally, while we support the mechanism outlined in Fig. 2, we cannot rule out the possibility of a radical chain pathway in which radical **11** abstracts an H-atom from alcohol **7** or thiol catalyst **5**.

In summary, this alkylation strategy represents the first, to our knowledge, general use of alcohols as simple alkylating agents and enables rapid late-stage derivatization of medicinally relevant molecules. Given the influence on drug pharmacokinetics and absorption, distribution, metabolism and excretion (ADME) properties, this method of installing inert alkyl groups will probably find wide application in the medicinal chemistry community. We have developed a mild and operationally simple alkylation reaction via the synergistic merger of photoredox and thiol hydrogen atom transfer organocatalysis to forge challenging heteroaryl C–C bonds using alcohols as latent nucleophiles. This bio-inspired strategy mimics the key step in enzyme-catalysed DNA biosynthesis via a new spin-centre shift elimination of H₂O to generate radical intermediates from simple alcohols.

Received 20 May; accepted 26 June 2015.

Published online 26 August 2015.

- Halliwell, B. & Gutteridge, J. M. C. *Free Radicals in Biology and Medicine* 4th edn (Oxford Univ. Press, 2007).

- Wessig, P. & Muehling, O. Spin-center shift (SCS) – a versatile concept in biological and synthetic chemistry. *Eur. J. Org. Chem.* 2219–2232 (2007).
- Minisci, F., Vismara, E. & Fontana, F. Homolytic alkylation of protonated heteroaromatic bases by alkyl iodides, hydrogen peroxide, and dimethyl sulfoxide. *J. Org. Chem.* **54**, 5224–5227 (1989).
- Molander, G. A., Colombel, V. & Braz, V. A. Direct alkylation of heteroaryls using potassium alkyl- and alkoxyethyltrifluoroborates. *Org. Lett.* **13**, 1852–1855 (2011).
- Ji, Y. *et al.* Innate C–H trifluoromethylation of heterocycles. *Proc. Natl Acad. Sci. USA* **108**, 14411–14415 (2011).
- Antonchick, A. P. & Burgmann, L. Direct selective oxidative cross-coupling of simple alkanes with heteroarenes. *Angew. Chem. Int. Edn Engl.* **52**, 3267–3271 (2013).
- DiRocco, D. A. *et al.* Late-stage functionalization of biologically active heterocycles through photoredox catalysis. *Angew. Chem. Int. Edn Engl.* **53**, 4802–4806 (2014).
- Eklund, H., Uhlir, U., Färnegårdh, M., Logan, D. T. & Nordlund, P. Structure and function of the radical enzyme ribonucleotide reductase. *Prog. Biophys. Mol. Biol.* **77**, 177–268 (2001).
- Schönherr, H. & Cernak, T. Profound methyl effects in drug discovery and a call for new C–H methylation reactions. *Angew. Chem. Int. Edn Engl.* **52**, 12256–12267 (2013).
- Minisci, F., Bernardi, R., Bertini, F., Galli, R. & Perchinunno, M. Nucleophilic character of alkyl radicals–VI: A new convenient selective alkylation of heteroaromatic bases. *Tetrahedron* **27**, 3575–3579 (1971).
- Duncun, M. A. J. Minisci reactions: versatile CH-functionalizations for medicinal chemists. *Med. Chem. Commun.* **2**, 1135–1161 (2011).
- Narayanan, J. M. R. & Stephenson, C. R. J. Visible light photoredox catalysis: applications in organic synthesis. *Chem. Soc. Rev.* **40**, 102–113 (2011).
- Prier, C. K., Rankic, D. A. & MacMillan, D. W. C. Visible light photoredox catalysis with transition metal complexes: applications in organic synthesis. *Chem. Rev.* **113**, 5322–5363 (2013).
- Schultz, D. M. & Yoon, T. P. Solar synthesis: prospects in visible light photocatalysis. *Science* **343**, 1239176 (2014).
- Qvortrup, K., Rankic, D. A. & MacMillan, D. W. C. A general strategy for organocatalytic activation of C–H bonds via photoredox catalysis: direct arylation of benzylic ethers. *J. Am. Chem. Soc.* **136**, 626–629 (2014).
- Hager, D. & MacMillan, D. W. C. Activation of C–H bonds via the merger of photoredox and organocatalysis: a coupling of benzylic ethers with Schiff bases. *J. Am. Chem. Soc.* **136**, 16986–16989 (2014).
- Cuthbertson, J. D. & MacMillan, D. W. C. The direct arylation of allylic sp³ C–H bonds via organic and photoredox catalysis. *Nature* **519**, 74–77 (2015).
- Jin, J. & MacMillan, D. W. C. Direct α -arylation of ethers through the combination of photoredox-mediated C–H functionalization and the Minisci reaction. *Angew. Chem. Int. Edn Engl.* **54**, 1565–1569 (2015).
- Ochiai, M. & Morita, K. A novel photo-induced methylation of pyrimidines and condensed pyrimidine compounds. *Tetrahedr. Lett.* **8**, 2349–2351 (1967).
- Stermitz, F. R., Wei, C. C. & Huang, W. H. Imine photoalkylations: quinoline and isoquinoline. *Chem. Commun. (Lond.)* **1968**, 482–483 (1968).
- Sugimori, A. *et al.* Radiation-induced alkylation of quinoline derivatives with alcohol. *Bull. Chem. Soc. Jpn.* **59**, 3905–3909 (1986).
- Slinker, J. D. *et al.* Efficient yellow electroluminescence from a single layer of a cyclometalated iridium complex. *J. Am. Chem. Soc.* **126**, 2763–2767 (2004).
- Shaidarova, L. G., Ziganshina, S. A. & Budnikov, G. K. Electrocatalytic oxidation of cysteine and cystine at a carbon-paste electrode modified with ruthenium(IV) oxide. *J. Anal. Chem.* **58**, 577–582 (2003).
- Escoubet, S. *et al.* Thiol radical mediated racemization of nonactivated aliphatic amines. *J. Org. Chem.* **71**, 7288–7292 (2006).
- Berkowitz, J., Ellison, G. B. & Gutman, D. Three methods to measure RH bond energies. *J. Phys. Chem.* **98**, 2744–2765 (1994).
- Roberts, B. P. Polarity-reversal catalysis of hydrogen-atom abstraction reactions: concepts and applications in organic chemistry. *Chem. Soc. Rev.* **28**, 25–35 (1999).
- Cai, Y. & Roberts, B. P. Radical-chain racemization of tetrahydrofurfuryl acetate under conditions of polarity-reversal catalysis: possible implications for the radical-induced strand cleavage of DNA. *Chem. Commun. (Camb.)* **1998**, 1145–1146 (1998).
- McNally, A., Prier, C. K. & MacMillan, D. W. C. Discovery of an α -amino C–H arylation reaction using the strategy of accelerated serendipity. *Science* **334**, 1114–1117 (2011).

Supplementary Information is available in the online version of the paper.

Acknowledgements Financial support was provided by NIHGM5 (R01 GM103558-03), and gifts from Merck and Amgen. J.J. thanks J. A. Terrett for assistance in preparing this manuscript.

Author Contributions J.J. performed and analysed experiments. J.J. and D.W.C.M. designed experiments to develop this reaction and probe its utility, and also prepared this manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.W.C.M. (dmacmill@princeton.edu).

Global separation of plant transpiration from groundwater and streamflow

Jaivime Evaristo¹, Scott Jasechko² & Jeffrey J. McDonnell^{1,3,4}

Current land surface models assume that groundwater, streamflow and plant transpiration are all sourced and mediated by the same well mixed water reservoir—the soil. However, recent work in Oregon¹ and Mexico² has shown evidence of ecohydrological separation, whereby different subsurface compartmentalized pools of water supply either plant transpiration fluxes or the combined fluxes of groundwater and streamflow. These findings have not yet been widely tested. Here we use hydrogen and oxygen isotopic data ($^2\text{H}/^1\text{H}$ ($\delta^2\text{H}$) and $^{18}\text{O}/^{16}\text{O}$ ($\delta^{18}\text{O}$)) from 47 globally distributed sites to show that ecohydrological separation is widespread across different biomes. Precipitation, stream water and groundwater from each site plot approximately along the $\delta^2\text{H}/\delta^{18}\text{O}$ slope of local precipitation inputs. But soil and plant xylem waters extracted from the 47 sites all plot below the local stream water and groundwater on the meteoric water line, suggesting that plants use soil water that does not itself contribute to groundwater recharge or streamflow. Our results further show that, at 80% of the sites, the precipitation that supplies groundwater recharge and streamflow is different from the water that supplies parts of soil water recharge and plant transpiration. The ubiquity of subsurface water compartmentalization found here, and the segregation of storm types relative to hydrological and ecological fluxes, may be used to improve numerical simulations of runoff generation, stream water transit time and evaporation–transpiration partitioning. Future land surface model parameterizations should be closely examined for how vegetation, groundwater recharge and streamflow are assumed to be coupled.

Freshwater fluxes via plant transpiration ($45,000 \text{ km}^3 \text{ yr}^{-1}$, ref. 3, to $62,000 \text{ km}^3 \text{ yr}^{-1}$, ref. 4), streamflow ($37,000 \text{ km}^3 \text{ yr}^{-1}$ to $40,000 \text{ km}^3 \text{ yr}^{-1}$, refs 5, 6) and groundwater recharge ($12,000 \text{ km}^3 \text{ yr}^{-1}$ to $16,200 \text{ km}^3 \text{ yr}^{-1}$, ref. 7) are central components of the terrestrial hydrosphere. Understanding the sources of water and processes that govern each component is important for predicting the effects of global change on water security and ecosystem services⁶. One of the most useful tools for quantifying water-cycle components and the linkages between plant ecology and physical hydrology is stable-isotope tracing⁸. Global isotopic databases developed over the past 60 years⁹ have enabled continental-scale assessments of transpiration/evaporation ratios⁴ and the recycling of rainfall back into the atmosphere¹⁰.

While global sets of precipitation⁹, streamflow⁹ and groundwater¹¹ data are now available for analysis, measurements of plant xylem

waters (that is, water moving within plants) remain dispersed throughout the primary, specialist literature. Synthesizing global groundwater, streamflow and plant xylem water isotopic data is important because recent watershed-based case studies have shown evidence of ecohydrological separation^{1,2}—meaning that the soil water that supplies plant transpiration is isolated from the water that recharges groundwater and replenishes streamflow. These two recent field studies both showed that plant transpiration is supplied by waters within unsaturated soils, but that local streamflow and groundwater were supplied by mobile water (linked to infiltrating precipitation) that moves through the soil seemingly unmixed with the waters that are retained in the soil.

Compartmentalization of a poorly mobile plant transpiration water pool versus a highly mobile stream/groundwater pool, if widespread, would challenge existing land surface model parameterizations that assume that plants and streams draw from a single, well mixed subsurface water reservoir¹². If true, such widespread ecohydrological separation would also have implications for isotope-based assessments of evaporation/transpiration ratios that rely on well mixed systems⁴. Here, we use a new global isotope database to test the ecohydrological compartmentalization hypothesis: that the isotopic composition of waters that supply plant transpiration differs from that of waters that supply groundwater and streamflow. The global ecohydrological isotope database consists of $^{18}\text{O}/^{16}\text{O}$ and $^2\text{H}/^1\text{H}$ ratios for plant xylem water ($n = 1,460$), soil water ($n = 1,830$), stream water ($n = 336$), groundwater ($n = 2,749$) and precipitation ($n = 488$) at 47 globally distributed locations (Table 1, Fig. 1).

Our approach is predicated on the knowledge that precipitation $\delta^2\text{H}$ and $\delta^{18}\text{O}$ values (see Methods for definitions) co-vary along a regression line with a $\delta^2\text{H}/\delta^{18}\text{O}$ slope of eight (this is the global meteoric water line, GMWL)¹³. The physical process of evaporation occurs under disequilibrium, produces a strong kinetic isotope effect that yields $\delta^2\text{H}/\delta^{18}\text{O}$ slopes of less than eight¹⁴, and results in a situation in which water samples that have undergone some evaporation plot ‘below’ the regression line of precipitation isotopic data. We use this well known difference between the meteoric water line and the local evaporation line as a key marker for ecohydrological compartmentalization^{1,2}.

Figure 1a–d shows isotopic data for groundwater, stream water, plant xylem water and soil water from our compiled database. Globally, headwater streams and groundwater plot approximately

Table 1 | Key information on 47 globally distributed isotopic data sets

Biome	Number of papers	RH (%)	MAT (°C)	MAP (mm yr ⁻¹)	LMWL slope	Plant $\delta^2\text{H}$ (‰)	Soil $\delta^2\text{H}$ (‰)	Stream $\delta^2\text{H}$ (‰)	GW $\delta^2\text{H}$ (‰)
Arid	7	49 ± 8.5	13 ± 5.2	314 (89)	8.0 (0.3)	−66 (39)	−44 (51)	−73 (15)	−27 (50)
Mediterranean	6	58 ± 7.3	15 ± 4.0	331 (157)	7.1 (2.5)	−48 (19)	−43 (27)	−46 (24)	−31 (17)
Temperate forests	17	58 ± 8.5	8.9 ± 5.0	533 (692)	8.2 (0.8)	−79 (36)	−79 (23)	−91 (48)	−84 (41)
Temperate grasslands	7	56 ± 5.1	16 ± 3.8	478 (662)	7.1 (0.5)	−28 (18)	−28 (10)	−22 (14)	−30 (41)
Tropics	10	65 ± 11	23 ± 3.8	1350 (1340)	8.2 (0.3)	−34 (33)	−38 (64)	−7.4 (30)	−14 (10)

RH, relative humidity; MAT, mean annual temperature; MAP, mean annual precipitation; LMWL, local meteoric water line; GW, groundwater. Values are mean ± 1 s.d. or median (interquartile range).

¹Global Institute for Water Security and School of Environment and Sustainability, University of Saskatchewan, Saskatoon, Saskatchewan S7N 3H5, Canada. ²Department of Geography, University of Calgary, Calgary, Alberta T2N 1N4, Canada. ³School of Geosciences, University of Aberdeen, Aberdeen AB34 3FX, UK. ⁴Department of Forest Engineering, Resources and Management, Oregon State University, Corvallis, Oregon 97331, USA.

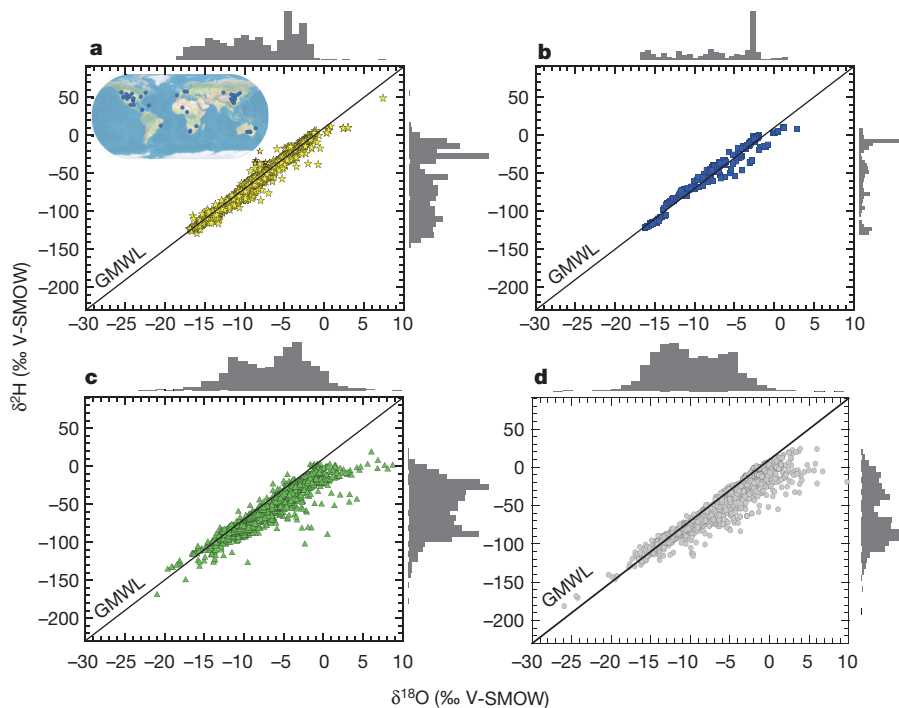


Figure 1 | $\delta^{18}\text{O}$ and $\delta^2\text{H}$ values of groundwater, stream water, plant xylem water and soil water at 47 globally distributed sites. The median (interquartile range) $\delta^{18}\text{O}$ and $\delta^2\text{H}$ values are: **a**, groundwater: -7.7 (7.4), -51.5 (62.6), $n = 2,749$; **b**, stream water: -6.2 (8.8), -37.1 (66.9), $n = 336$; **c**, plant xylem water: -5.5 (6.1), -50.6 (50.6), $n = 1,460$; **d**, soil water: -7.5 (7.4), -63.9 (52.2), $n = 1,830$. The inset in **a** shows the locations of 47 globally distributed stable isotopic data sets. The histogram borders show partitioning of the data sets at 30 identical intervals or bins. The global meteoric water line (GMWL)¹³ is also shown. V-SMOW, Vienna-standard mean ocean water.

along the GMWL. These patterns suggest that stream water and groundwater follow the local precipitation input signal¹⁵. Plant xylem and soil waters extracted from the 47 studies plot below the regression of global meteoric waters—a result of the strong kinetic isotope effect via the process of evaporation¹⁴.

To quantify the similarities or differences between waters used by plants and waters that contribute to groundwater and streamflow, we use a site-by-site comparison based on a precipitation offset¹⁶:

$$\text{Precipitation offset} = [\delta^2\text{H} - a \delta^{18}\text{O} - b] / S \quad (1)$$

where a and b are the slope and y intercept, respectively, calculated from monthly measurements of $\delta^{18}\text{O}$ and $\delta^2\text{H}$ from local precipitation at each study site, and S is one standard deviation measurement uncertainty for both $\delta^{18}\text{O}$ and $\delta^2\text{H}$. The precipitation offset describes the difference in the isotopic composition of environmental waters from that of local precipitation, which has, by definition, a precipitation offset of zero. The precipitation offset can distinguish hydrological processes that occur under chemical equilibrium (for example, the

condensation of vapour¹³) from hydrological processes that occur under disequilibrium (for example, evaporation¹⁷). Plant transpiration does not affect the precipitation offset, whereas the evaporation of meteoric water near the land surface results in precipitation offset values of less than zero. By comparing the local precipitation offsets of our four water types (that is, soil water, plant xylem water, stream water and groundwater), we can use the stable isotopes to distinguish evaporated waters from non-evaporated waters and to test whether streamflow, groundwater and plant transpiration are supplied by one well mixed subsurface water reservoir, or more than one water reservoir (namely water that is retained in the soil and water that recharges groundwater and discharges in streams).

Figure 2 shows that plant xylem water offsets (median, interquartile range, $P < 0.0001$ using nonparametric Steel–Dwass method) (-5.6 , 4.7) and soil water offsets (-6.2 , 4.4) are significantly different from the offsets of groundwater (-1.8 , 3.2) and stream water (0.22 , 3.7) in all five of the biomes represented by the 47 sites in our database. Of our 47 sites, 40 have groundwater precipitation offsets that are

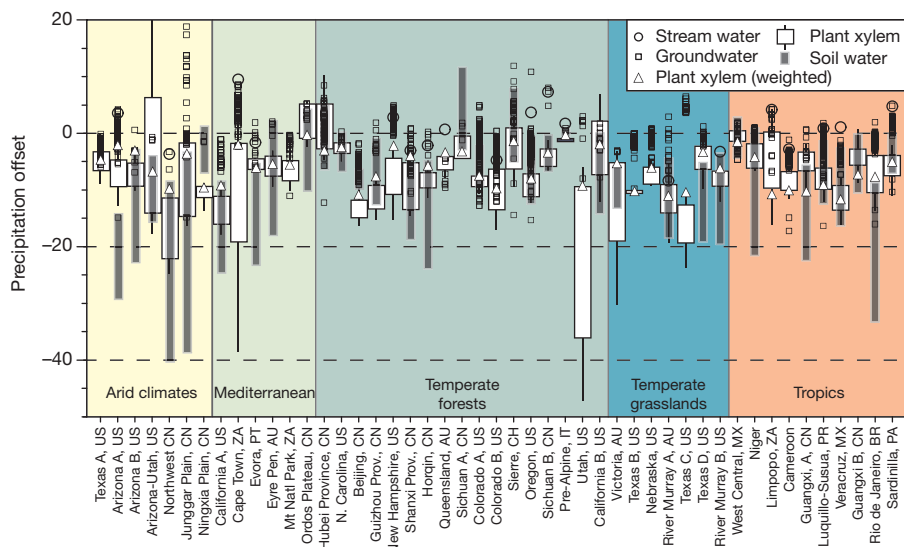


Figure 2 | Precipitation offset values of groundwater, stream water, plant xylem water and soil water for 47 sites grouped by biome. Extents of plant xylem (white) and soil (grey) water bars show 25th and 75th percentiles. All values of groundwater (squares) are shown for visualization of data density (that is, darker regions) and dispersion (that is, lighter regions). Mean values of stream water (circles) are also shown, as are the transpiration-amount-weighted values of plant xylem water (triangles).

statistically distinct ($P < 0.05$ using two-tailed homoscedastic/heteroscedastic tests, as applicable) from both soil water and plant xylem water precipitation offsets. Our analysis is suggestive of a widespread occurrence of ecohydrological separation—that is, poor and incomplete mixing of subsurface water, with one reservoir of water sustaining plant transpiration, and another contributing to groundwater recharge and streamflow. On a site-by-site basis, groundwater and stream water have a precipitation offset that is on average respectively 5.4 and 4.8 higher (that is, closer to zero) than do soil and plant xylem waters. The greatest differences between the precipitation offsets of streamwater/groundwater and plant xylem/soil water are found in the tropical and Mediterranean biomes (7.7 and 5.4, respectively), with smaller differences observed in the arid, temperate grassland, and temperate forest biomes (3.6, 2.4 and 1.6 on average, respectively).

Recent work has shown that different storm types contribute disproportionately to groundwater recharge (see, for example, refs 11, 18). Some studies have shown that more intense storms dominate groundwater recharge¹⁸; others present evidence to the contrary¹⁹. While our analyses do not allow us to associate storm intensity with either plant transpiration or groundwater recharge fluxes, we can nevertheless trace the isotopic composition of the precipitation from which plant xylem water originated. We calculated the intersection points of local plant xylem evaporation lines with local meteoric water lines (LMWLs)—that is, plant xylem δ source value (see Extended Data Fig. 1 and Methods):

$$\delta^2\text{H intercept} = \delta^2\text{H} - m \delta^{18}\text{O} \quad (2)$$

$$\delta^{18}\text{O intercept} = [\delta^2\text{H intercept} - b]/a \quad (3)$$

where m , a and b are the slope of the evaporation line, the LMWL slope, and the LMWL intercept, respectively.

The results of this analysis show that at 80% of the sites (see Extended Data Table 1; 37 of 46 sites) where plant xylem water δ source values can be calculated, groundwater isotope values (median, interquartile range, $P < 0.05$ using nonparametric Wilcoxon method) (-52 , 63‰ $\delta^2\text{H}$) are statistically different from plant xylem water δ source values (-82 , 83‰ $\delta^2\text{H}$). This suggests that, in many cases (see, for example, Extended Data Fig. 2), ecologically and hydrologically important precipitation is segregated in both space and time, even before these waters become further segregated in the subsurface for plant transpiration or for groundwater recharge and streamflow (see Methods and Extended Data Figs 2 and 4).

We also use equations (2) and (3) to trace the isotopic composition of precipitation from which soil water originated—that is, the soil water δ source value. We find that at 83% of the sites (Extended Data Table 1; 29 of 35 sites) where soil water isotopic data are available, soil water δ source values (-104 , 96‰ $\delta^2\text{H}$) are statistically different from groundwater isotope values. The significant difference between soil water δ source and groundwater isotope values suggests that some forms of precipitation that recharge the subsurface may be more important than others to plant transpiration fluxes. We assess the uncertainties in parameter m (equation (2)) and find overall average uncertainties of 1.07‰ for $\delta^{18}\text{O}$ and 5.54‰ for $\delta^2\text{H}$ (2σ). These are slightly less than, but somewhat comparable to, the prediction uncertainties in precipitation isotope values (1.17‰ for $\delta^{18}\text{O}$ and 9.4‰ for $\delta^2\text{H}$; ref. 20).

Plants regulate water fluxes from the subsurface to the atmosphere⁴. Our discovery that ecohydrological separation is widespread throughout the terrestrial water cycle has major implications for isotope-based estimates of runoff sources¹², streamwater residence times²¹ and evaporation/transpiration partitioning⁴. Recent estimates⁴ of catchment-scale transpiration/evapotranspiration (T/ET) ratios have followed an assumption of well mixed water stores within the critical zone, consistent with most land surface parameterizations¹²; our findings fundamentally challenge this assumption as it relates to

catchment-based evapotranspiration partitioning^{4,22,23} and most land surface models¹². Our work would suggest that downstream water isotope compositions are biased towards precipitation and groundwater source contributions, and do not reflect the composition of water seen in soil. This in turn casts doubt on the estimates of transpiration/evapotranspiration made in other studies if based solely on isotope data, meaning that evapotranspiration partitioning based on downstream water isotope compositions may not represent an integrated catchment-wide isotopic signature as widely applied.

Notwithstanding these issues, our general finding that transpiration comprises the greatest fraction of terrestrial evapotranspiration is reinforced by the lines of evidence discussed in ref. 4, and by the results of land surface models (terrestrial T/ET of 59% to 80%; refs 24, 25), atmospheric vapour isotope measurements (European T/ET of 62%; ref. 26), global syntheses of stand-level transpiration measurements (terrestrial T/ET of roughly 61%; ref. 3), and some but not all general circulation models (see refs 27, 28). Although transpiration is, indeed, the largest component of terrestrial evapotranspiration⁴, our results show that the mechanisms by which such partitioning takes place, and links to other components of the water cycle²⁹, are still poorly understood. These combined findings point the way towards the research that is needed to understand the ecophysiological basis of ecohydrological separation across biomes. Finally, our results also suggest that existing land surface model parameterizations of plant physiological processes and runoff³⁰ (that is, streamflow) can be made more realistic through the incorporation of ecohydrological separation.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 27 September 2014; accepted 14 July 2015.

- Brooks, J. R., Barnard, H. R., Coulombe, R. & McDonnell, J. J. Ecohydrologic separation of water between trees and streams in a Mediterranean climate. *Nature Geosci.* **3**, 100–104 (2010).
- Goldsmith, G. R. *et al.* Stable isotopes reveal linkages among ecohydrological processes in a seasonally dry tropical montane cloud forest. *Ecohydrol.* **5**, 779–790 (2012).
- Schlesinger, W. H. & Jasechko, S. Transpiration in the global water cycle. *Agric. For. Meteorol.* **189**, 115–117 (2014).
- Jasechko, S. *et al.* Terrestrial water fluxes dominated by transpiration. *Nature* **496**, 347–350 (2013).
- Dai, A. & Trenberth, K. E. Estimates of freshwater discharge from continents: latitudinal and seasonal variations. *J. Hydrometeorol.* **3**, 660–687 (2002).
- Oki, T. & Kanae, S. Global hydrological cycles and world water resources. *Science* **313**, 1068–1072 (2006).
- Wada, Y., Van Beek, L. P. H., Wanders, N. & Bierkens, M. F. P. Human water consumption intensifies hydrological drought worldwide. *Environ. Res. Lett.* **8**, 034036 (2013).
- Yakir, D. & Wang, X.-F. Fluxes of CO₂ and water between terrestrial vegetation and the atmosphere estimated from isotope measurements. *Nature* **380**, 515–517 (1996).
- International Atomic Energy Agency's Water Resources Programme <http://www.iaea.org/water/> (2014).
- Levin, N. E., Zipser, E. J. & Cerling, T. E. Isotopic composition of waters from Ethiopia and Kenya: Insights into moisture sources for eastern Africa. *J. Geophys. Res.* **D 114**, D23306 (2009).
- Jasechko, S. *et al.* The pronounced seasonality of global groundwater recharge. *Wat. Resour. Res.* **50**, 8845–8867 (2014).
- Birkel, C., Tetzlaff, D., Dunn, S. M. & Soulsby, C. Towards a simple dynamic process conceptualization in rainfall-runoff models using multi-criteria calibration and tracers in temperate, upland catchments. *Hydrol. Processes* **24**, 260–275 (2010).
- Friedman, I. Deuterium content of natural waters and other substances. *Geochim. Cosmochim. Acta* **4**, 89–103 (1953).
- Craig, H. Isotopic variations in meteoric waters. *Science* **133**, 1702–1703 (1961).
- Dutton, A. R. Groundwater isotopic evidence for paleorecharge in U.S. High Plains aquifers. *Quat. Res.* **43**, 221–231 (1995).
- Landwehr, J. & Coplen, T. in *Isotopes in Environmental Studies* 132–135 (IAEA-CN-118/56, International Atomic Energy Agency, 2006).
- Dansgaard, W. Stable isotopes in precipitation. *Tellus* **16**, 436–468 (1964).
- Taylor, R. G. *et al.* Evidence of the dependence of groundwater resources on extreme rainfall in East Africa. *Nature Clim. Change* **3**, 374–378 (2013).
- Scholl, M. A. & Murphy, S. F. Precipitation isotopes link regional climate patterns to water supply in a tropical mountain forest, eastern Puerto Rico. *Wat. Resour. Res.* **50**, 4305–4322 (2014).

20. Good, S. P. *et al.* Patterns of local and nonlocal water resource use across the western US determined via stable isotope intercomparisons. *Wat. Resour. Res.* **50**, 8034–8049 (2014).
21. Syed, T. H., Famiglietti, J. S., Zlotnicki, V. & Rodell, M. Contemporary estimates of Pan-Arctic freshwater discharge from GRACE and reanalysis. *Geophys. Res. Lett.* **34**, L19404 (2007).
22. Ferguson, P. R., Weinrauch, N., Wassenaar, L. I., Mayer, B. & Veizer, J. Isotope constraints on water, carbon, and heat fluxes from the northern Great Plains region of North America. *Glob. Biogeochem. Cycles* **21**, GB2023 (2007).
23. Gibson, J. J. & Edwards, T. W. D. Regional water balance trends and evaporation-transpiration partitioning from a stable isotope survey of lakes in northern Canada. *Glob. Biogeochem. Cycles* **16**, 10-1-10-14 (2002).
24. Dirmeyer, P. A. *et al.* GSWP-2: multimodel analysis and implications for our perceptions of the land surface. *Bull. Am. Meteorol. Soc.* **87**, 1381–1397 (2006).
25. Wang-Erlandsson, L., van der Ent, R. J., Gordon, L. J. & Savenije, H. H. G. Contrasting roles of interception and transpiration in the hydrological cycle—part 1: temporal characteristics over land. *Earth Syst. Dyn.* **5**, 441–469 (2014).
26. Aemisegger, F. *et al.* Deuterium excess as a proxy for continental moisture recycling and plant transpiration. *Atmos. Chem. Phys.* **14**, 4029–4054 (2014).
27. Sutanto, S. J. *et al.* A perspective on isotope versus non-isotope approaches to determine the contribution of transpiration to total evaporation. *Hydrol. Earth Syst. Sci.* **18**, 2815–2827 (2014).
28. Lawrence, D. M., Thornton, P. E., Oleson, K. W. & Bonan, G. B. The partitioning of evapotranspiration into transpiration, soil evaporation, and canopy evaporation in a GCM: Impacts on land-atmosphere interaction. *J. Hydrometeorol.* **8**, 862–880 (2007).
29. Gouet-Kaplan, M., Tartakovsky, A. & Berkowitz, B. Simulation of the interplay between resident and infiltrating water in partially saturated porous media. *Wat. Resour. Res.* **45**, W05416 (2009).
30. Stöckli, R., Vidale, P. L., Boone, A. & Schär, C. Impact of scale and aggregation on the terrestrial water exchange: integrating land surface models and Rhône catchment observations. *J. Hydrometeorol.* **8**, 1002–1015 (2007).

Acknowledgements J.E. thanks the Saskatchewan Innovation and Opportunity Scholarship, Global Institute for Water Security, and School of Environment and Sustainability (University of Saskatchewan) for financial support.

Author Contributions J.J.M. conceived the idea of testing the ecohydrological compartmentalization hypothesis with global data. J.E., S.J. and J.J.M. brainstormed on how to do this. J.E. designed the approach, compiled the data set, and conducted the statistical analyses. J.E. wrote the first paper draft. S.J. and J.J.M. edited and commented on the manuscript and contributed to the text in later iterations.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.E. (jaivime.evaristo@usask.ca).

METHODS

Data compilation and treatment. We performed a keyword-based search of published literature for stable water isotopes in ecology and hydrology. Because ecohydrological separation³¹ is based on the offset of a water sample from the local meteoric water line (that is, 'precipitation offset'¹⁶; equation (1)), we included only dual-isotope findings and excluded papers that used either $\delta^2\text{H}$ or $\delta^{18}\text{O}$ alone. Stable isotope values from the 47 papers found were then extracted in one of two ways: first, where data were reported in tabular form, we compiled the data directly into the database; second, where plant xylem and soil water isotope data were not reported in tabular form, we used a graphical user interface to extract data points from figures in the original paper. We then calculated the precipitation offset values on the basis of equation (1). The measurement uncertainty S in equation (1) was calculated as:

$$S = \left[(\delta^2\text{H analytical error})^2 + (\delta^{18}\text{O analytical error})^2 \right]^{0.5} \quad (4)$$

Reported analytical errors for $\delta^2\text{H}$ and $\delta^{18}\text{O}$ are 1‰ and 0.2‰ on average, respectively.

We extracted groundwater isotope data for 45 of 47 sites either from the compiled papers ($n = 24$) or from the comprehensive global groundwater database ($n = 21$) of ref. 11. Of the 21 groundwater data sets compiled using the latter database, 16, 2, 1 and 2 data sets are within a 200-, 300-, 400- and 500-km radius of actual study sites. The radii within which groundwater data were extracted were chosen so that we could build groundwater data sets for most of the 47 sites in our database. To test whether or not the choice of radii imposed a scale-dependent variation (that is, bias) in isotopic trends, we performed a sensitivity analysis by calculating the precipitation offset values of groundwater at distances of 25, 50 and 100 km. We found that precipitation offset values of groundwater did not differ statistically in space. That is, precipitation offset of groundwater at 25 km (-3.5 ± 2.2 , $n = 688$) was not statistically different from precipitation offset at 50 km (-2.5 ± 2.4 , $n = 1,605$), 100 km (-2.4 ± 2.4 , $n = 3,295$), 200 km (-2.7 ± 2.2 , $n = 6,598$), 300 km (-2.5 ± 4.5 , $n = 12,000$), 400 km (-2.8 ± 4.6 , $n = 18,239$) and 500 km (-2.8 ± 4.8 , $n = 24,000$). This scale-invariant behaviour of groundwater precipitation offset supported our choice of radii in building the data sets for 45 of 47 sites in our database. It also reinforced one of the key messages of this work, in that groundwater isotopes generally fall along the local meteoric water line.

To show that plant transpiration water and groundwater recharge are related to different storm types, we traced the precipitation δ source value of plant xylem water by calculating the intersection points of local evaporation lines with local meteoric water lines (LMWLs) (equations (2) and (3); Extended Data Fig. 1). On a site-by-site basis, we compared the calculated precipitation δ source value of plant xylem water and soil water with the mean groundwater δ value (Extended Data Table 1).

Comparing plant xylem water δ source values with mean groundwater δ values requires intuitively that both should be situated as close to each other as possible at a site. The distance of groundwater wells to actual study sites in our database, however, varies from 0 km to almost 500 km. To test whether our approach of comparing both isotope composition values was statistically robust, we ran a sensitivity analysis by comparing plant xylem water δ source values with only the closest groundwater well to a given site. Increasing the radii between actual study sites and sites of groundwater measurements was then used as a critical evaluation metric for the approach (Extended Data Fig. 3). Our results showed that, for five increasing radii ranges between the actual xylem water study site and groundwater well site, the differences (median (interquartile ranges), absolute $\delta^2\text{H}$ ‰) between plant xylem water δ and groundwater δ values (24 (29), $n = 7$; 30 (30), $n = 8$; 31 (42), $n = 7$; 21 (22), $n = 9$; 23 (40), $n = 11$) are not statistically different from each other ($P > 0.90$, Tukey–Kramer honest significant difference). This suggests that our approach in comparing plant xylem water δ source values (that is, xylem evaporation line intercept with LMWL) and mean groundwater value at a site is valid. We underline that this does not imply that groundwater isotope values are invariant in space, but rather that the mean difference between plant xylem water δ source values and mean groundwater values is invariant in space (statistically not different), as shown in Extended Data Fig. 3.

We make a distinction between the two phenomena: 'segregation' of storm types and 'ecohydrological separation'. The former is related to source precipitation analysis (equations (2) and (3)), the latter to the fate of these waters either as groundwater or for plant transpiration (equation (1)). Segregation of storm types and ecohydrological separation in space is ubiquitous in the global data set. We are unable to test for both phenomena in time because of limitations in the available information in the compiled source papers. That is, if a source paper has data for at least two time points (usually contrasting moisture time points) then we can use such information to explore temporal contrasts (38 of 47 sites). For the 38 sites

that satisfy this criterion, both storm-type segregation and ecohydrological separation exist in 30 and 32 of 38 sites, respectively ($P < 0.05$ using nonparametric Wilcoxon Method).

We recognize that non-weighted plant xylem water isotope values would be biased towards values where transpiration rates are low. To test the robustness of the precipitation offset parameter, we also calculate the transpiration-amount-weighted isotopic composition of plant xylem water ($\delta_{\text{xyl}(\text{weighted})}$) using compiled long-term, global, biome-level transpiration rate estimates³:

$$\delta_{\text{xyl}(\text{weighted})} = \frac{\sum_{i=1}^n \delta_{\text{xyl}(i)} T_i}{\sum_{i=1}^n T_i} \quad (5)$$

where $\delta_{\text{xyl}(i)}$ represents the isotopic composition of xylem water during sampling month i , and T_i represents the amount of transpiration during month i . As illustrated in Fig. 2, both transpiration-amount-weighted and non-weighted plant xylem precipitation offsets are statistically different from zero, supporting our primary conclusion that plant transpiration water chemistries are different from groundwater and streamflow at 40 of 47 locations. We use no amount weighting on groundwater isotope values, in agreement with observations that showed little change in groundwater isotopic composition on timescales of years and decades^{32,33}.

To trace the fate of water after precipitation (that is, either as groundwater recharge or as plant water uptake), we quantified the precipitation offset from the LMWL (equation (1)). We confirmed ecohydrological separation at a study site if plant xylem water and soil water isotopic composition fall below the regression of $\delta^2\text{H}$ and $\delta^{18}\text{O}$ values in local precipitation on the LMWL.

Conventional notation for isotope composition is used where $\delta = (R_{\text{sample}}/R_{\text{standard}} - 1) \times 1,000\text{‰}$, where R is the ratio of $^{18}\text{O}/^{16}\text{O}$ ($\delta^{18}\text{O}$) or $^2\text{H}/^1\text{H}$ ($\delta^2\text{H}$) in the sample, or in the international standard (Vienna-standard mean ocean water, V-SMOW).

Statistical analysis. Parametric requirements of normality and equal variances, particularly for aggregate precipitation offset values, are not satisfied via attempts to transform the data. Testing whether group means are located similarly across groups is performed using nonparametric tests, which use functions of the response ranks (or rank scores). A Kruskal–Wallis/Steel–Dwass method is performed to test whether or not the precipitation offset values of the water types—groundwater, stream water, plant xylem water and soil water—differ statistically from each other. We perform a similar nonparametric test (Dunn all pairs for joint ranks method) by computing ranks on all the data. The results are the same as those from the pairwise method Kruskal–Wallis/Steel–Dwass test. To test whether each water type is statistically different from zero (that is, the precipitation offset value of local precipitation), the Dunn method for joint ranking is performed. The test shows that plant xylem water and soil water are statistically different from zero, while groundwater and stream water are not statistically different from zero. This test result supports the interpretation that groundwater and stream water fall along the $\delta^2\text{H}/\delta^{18}\text{O}$ slopes of local meteoric water lines, while plant xylem water and soil water fall 'below' the slopes of this linear regression. The same method is also used to test for statistical significance of precipitation offset values of each water type across biomes. These nonparametric tests are based on ranks and control for the overall alpha level ($\alpha = 0.05$). The Dunn method, which reports P values after a Bonferroni adjustment, is used to correct for multiple testing problem that may arise from an inflated type I error rate ($0.0001 \leq P \leq 0.05$). Where parametric requirements are met, particularly for intrasite tests on water types, Student's t /Tukey–Kramer HSD tests are performed as applicable. Uncertainty estimation, particularly for equations (2) and (3) parameters, is performed with the jack-knifing approach³⁴.

A mechanism for ecohydrological separation. Partial mixing of 'new' (incoming) and 'old' (resident) water in the subsurface is rarely considered in conceptual models^{35,36}. Our key finding that groundwater/stream water and soil/plant uptake water are fundamentally (physically and temporally) separated supports the dynamic partial mixing model of ref. 37. In fact, it was the contrasting conclusions drawn by ref. 1 compared with those of refs 38 and 39 regarding the mixing mechanisms that led the authors of ref. 37 to propose the use of the following dimensionless mixing coefficient $C_{M,i}$, controlled mainly by soil moisture content:

$$C_{M,i} = \frac{1}{2} - \frac{1}{2} \operatorname{erf} \left(\frac{S_U}{S_{U_{\max}}} - \mu_{C_{M,i}} \right) / \left(\frac{\sigma_{C_{M,i}}}{\sqrt{2}} \right) \quad (6)$$

where S_U and $S_{U_{\max}}$ are actual storage and storage capacity within the root zone, respectively; $\mu_{C_{M,i}}$ and $\sigma_{C_{M,i}}$ are location and shape parameters, respectively; and i is the storage compartment. Equation (6) is applied to tracer (for example, stable water isotopes) balance equations, which may then enable functional comparisons amongst other alternative diagnostic models (for example, the more widely used complete mixing model).

Our precipitation offset parameter analysis (equation (1)) is used to modify equation (6) by substituting the precipitation offset value of soil water for the term $\frac{S_U}{S_{U_{\max}}}$:

$$C_{M,i} = \frac{1}{2} - \frac{1}{2} \operatorname{erf} \left(\frac{|P_x| - \mu C_{M,i}}{\sigma C_{M,i} \sqrt{2}} \right) \quad (7)$$

where $|P_x|$ is the absolute value of precipitation offset parameter. This results in a dimensionless mixing coefficient $C_{M,i}$ value that decreases as precipitation offset $|P_x|$ value increases. When $C_{M,i}$ is applied in tracer mass balance equations (as outlined in ref. 37), mixing between 'new' and 'old' water increases as soil moisture decreases; or, conversely, separation between 'new', 'fast-flowing' waters and 'old', 'matrix' waters increases with higher antecedent soil moisture. The persistence of 'old' water within the soil matrix and reduced participation in dispersive and diffusive exchange with preferential flow path water lead to continued exposure to evaporation (stage 1, capillary action, and stage 2, vapour diffusion). For details regarding evaporation from porous media, see ref. 40.

Our conceptual formulation as outlined in equation (7) is supported by the results of our precipitation offset analysis. Our analysis provides a site-by-site (Extended Data Table 2) and biome-level (Extended Data Table 3) quantification of the magnitude of separation—and, by extension, mixing—between ground-water recharge and stream discharge, and the water that recharges the soil matrix and is being taken up by plants for transpiration. Extended Data Table 3 shows that in soils of the arid biome, the precipitation offset value is highest (that is, closer to zero); conversely, in soils of the humid tropics where antecedent soil wetness is high, the precipitation offset value is lower. Calculating the dimensionless mixing coefficient $C_{M,i}$ using the precipitation offset values in Extended Data Table 3 and plugging these values into equation (7) supports the observation that in the dry soils of the arid biome, mixing between new, fast-flowing waters and old, matrix waters increases. The opposite is true for the other extreme, in humid tropical soils, where antecedent soil wetness is high. In general, because plants in our compiled database use soil water, these precipitation offset trends in soils are therefore consistent with plant xylem water data. That is, the magnitude of ecohydrological separation—plants using evaporated soil water that is isotopically distinct from groundwater recharge and stream discharge—increases with antecedent soil wetness. The relationship between soil wetness and the dimensionless mixing coefficient $C_{M,i}$ is discussed in detail and tested with actual, long-term catchment-level data in ref. 37. However, we state a caveat: the use of the precipitation offset parameter in equation (7) may be considered as a coarse (first-order) approximation given the nonlinear relationship between evaporative loss and the precipitation offset parameter.

While ref. 1 was the first paper to develop the ecohydrological separation concept and was relatively successful at proposing a mechanistic explanation for the observed results, other work has shown that such a mechanism may not universally explain the observed ecohydrological separation. For example, ref. 2 also found ecohydrological separation in a seasonally dry cloud forest in Mexico; these authors argued that the mechanism proposed in ref. 1 was not likely to explain the observed isotopic separation in their study². The plant xylem water values in ref. 1 are more enriched than most of the soil water values—the opposite case to ref. 2. If the 'first in, last out' mechanism proposed by ref. 1 was correct, then the measured plant xylem values should have matched those of (or at least be bounded by) the measured soil water values. Their data suggest that this was not the case. In contrast, the authors of ref. 2 observed their plant xylem water values to lie completely in between precipitation and bulk soil water values. The aggregate result (Extended Data Fig. 4) from our global data set lends support more to the interpretation of ref. 2 than to that of ref. 1.

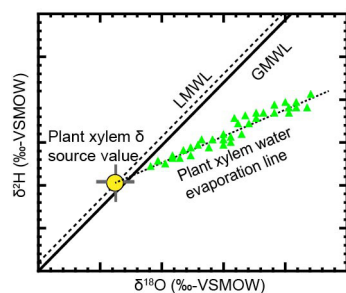
Water extraction techniques. As underlined in our central message, plant xylem water and soil water isotopes plot 'off' the LMWLs, supporting the idea of a widespread occurrence of ecohydrological separation on a global scale. This finding is true across the different techniques used to extract water out of soil and plant stem samples in our data set. The authors of ref. 1 argued that plant transpiration is supplied by 'tightly bound' waters within unsaturated soils. This interpretation was inferred from the laboratory technique used to extract water out of a soil sample (cryogenic vacuum distillation), which uses suction pressures that are orders of magnitude greater than those used in other field techniques (for example, suction lysimetry). Potential nuances in the fidelity of water extraction from soil samples using existing laboratory techniques have recently been explored^{41–43}. These findings suggest that soil physicochemical characteristics may contribute to isotopic fractionation, specifically with respect to $\delta^{18}\text{O}$. We explored the relationship between water extraction techniques and plant xylem water/soil water $\delta^{18}\text{O}$ in our data set. Extended Data Fig. 5 shows the plant xylem water/soil water $\delta^{18}\text{O}$ values using a liquid–vapour equilibration technique from cryogenic vacuum

distillation and azeotropic distillation. Although there are statistically significant differences ($P < 0.0001$, nonparametric Dunn method for joint ranking) between both cryogenic vacuum ($n = 2,640$) and azeotropic distillation ($n = 441$), and liquid–vapour equilibration methods ($n = 204$), there is no significant difference in plant xylem water $\delta^{18}\text{O}$ between the two more widely used techniques, cryogenic vacuum and azeotropic distillation ($P = 0.35$, nonparametric Dunn method for joint ranking). Despite these differences in $\delta^{18}\text{O}$ of plant xylem water and soil water with respect to water extraction techniques, both water types plot 'off' the LMWL in dual-isotope space. This suggests that ecohydrological separation exists beyond any differences in soil water $\delta^{18}\text{O}$ that are related to different water extraction techniques.

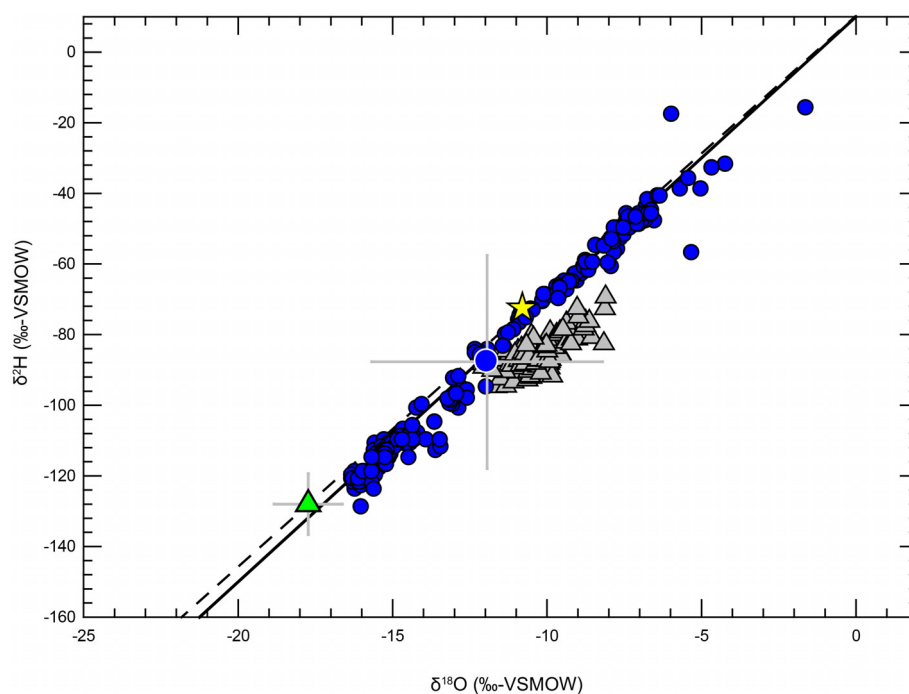
Global map of plant xylem water $\delta^2\text{H}$ and $\delta^{18}\text{O}$. For the first time, to our knowledge, we provide not only a global map of plant xylem $\delta^2\text{H}$ and $\delta^{18}\text{O}$, but also their relationship to respective LMWLs as integrated in the precipitation offset parameter—a fundamental descriptor of ecohydrological separation (Extended Data Fig. 6). Our compilation of global plant xylem $\delta^2\text{H}$ and $\delta^{18}\text{O}$ may complement other existing large-scale isotopic data sets from precipitation⁴⁴ and streams⁴⁵, in pursuing future research questions related to plant–water relations from continental to global scales.

- McDonnell, J. J. The two water worlds hypothesis: ecohydrological separation of water between streams and trees? *WIREs Water* **1**, 323–329 (2014).
- Darling, W. G., Bath, A. H. & Talbot, J. C. The O and H stable isotopic composition of fresh waters in the British Isles. 2. Surface waters and groundwater. *Hydrol. Earth Syst. Sci.* **7**, 183–195 (2003).
- Genty, D. *et al.* Rainfall and cave water isotopic relationships in two South-France sites. *Geochim. Cosmochim. Acta* **131**, 323–343 (2014).
- Wu, C. Jackknife, bootstrap and other resampling methods in regression analysis—discussion. *Ann. Stat.* **14**, 1261–1295 (1986).
- van der Velde, Y., Torfs, P. J. J. F., van der Zee, S. E. A. T. M. & Uijlenhoet, R. Quantifying catchment-scale mixing and its effects on time-varying travel time distributions. *Wat. Resour. Res.* **48**, W06536 (2012).
- Page, T., Beven, K. J., Freer, J. & Neal, C. Modelling the chloride signal at Plynlion, Wales, using a modified dynamic TOPMODEL incorporating conservative chemical mixing (with uncertainty). *Hydrol. Processes* **21**, 292–307 (2007).
- Hrachowitz, M., Savenije, H., Bogaard, T. A., Tetzlaff, D. & Soulsby, C. What can flux tracking teach us about water age distribution patterns and their temporal dynamics? *Hydrol. Earth Syst. Sci.* **17**, 533–564 (2013).
- Legout, C. *et al.* Solute transfer in the unsaturated zone—groundwater continuum of a headwater catchment. *J. Hydrol.* **332**, 427–441 (2007).
- Klaus, J., Zehe, E., Elsner, M., Külls, C. & McDonnell, J. J. Macropore flow of old water revisited: experimental insights from a tile-drained hillslope. *Hydrol. Earth Syst. Sci.* **17**, 103–118 (2013).
- Or, D., Lehmann, P., Shahraeeni, E. & Shokri, N. Advances in soil evaporation physics—a review. *Vadose Zone J.* **12**(4), <http://dx.doi.org/10.2136/vzj2012.0163> (2013).
- Oerter, E. *et al.* Oxygen isotope fractionation effects in soil water via interaction with cations (Mg, Ca, K, Na) adsorbed to phyllosilicate clay minerals. *J. Hydrol.* **515**, 1–9 (2014).
- Meissner, M., Koehler, M., Schwendenmann, L., Hoelscher, D. & Dyckmans, J. Soil water uptake by trees using water stable isotopes ($\delta^2\text{H}$ and $\delta^{18}\text{O}$)—a method test regarding soil moisture, texture and carbonate. *Plant Soil* **376**, 327–335 (2014).
- Orlowski, N., Frede, H., Brüggemann, N. & Breuer, L. Validation and application of a cryogenic vacuum extraction system for soil and plant water extraction for isotope analysis. *J. Sensors Sensor Syst.* **2**, 179–193 (2013).
- Bowen, G. J. & Wilkinson, B. Spatial distribution of $\delta^{18}\text{O}$ in meteoric precipitation. *Geology* **30**, 315–318 (2002).
- Kendall, C. & Coplen, T. B. Distribution of oxygen-18 and deuterium in river waters across the United States. *Hydrol. Processes* **15**, 1363–1393 (2001).
- Boutton, T. W., Archer, S. R. & Midwood, A. J. Stable isotopes in ecosystem science: structure, function and dynamics of a subtropical savanna. *Rapid Commun. Mass Spectrom.* **13**, 1263–1277 (1999).
- McKeon, C. *et al.* Growth and water and nitrate uptake patterns of grazed and ungrazed desert shrubs growing over a nitrate contamination plume. *J. Arid Environ.* **64**, 1–21 (2006).
- Snyder, K. A. & Williams, D. G. Water sources used by riparian trees varies among stream types on the San Pedro River, Arizona. *Agric. For. Meteorol.* **105**, 227–240 (2000).
- Williams, D. G. & Ehleringer, J. R. Intra- and interspecific variation for summer precipitation use in pinyon-juniper woodlands. *Ecol. Monogr.* **70**, 517–537 (2000).
- Zhou, Y.-D., Chen, S.-P., Song, W.-M., Lu, Q. & Lin, G.-H. Water-use strategies of two desert plants along a precipitation gradient in northwestern China. *Chinese J. Plant Ecol.* **35**, 789–800 (2011).
- Hai, Z., Xin-Jun, Z., Li-Song, T. & Yan, L. Differences and similarities between water sources of *Tamarix ramosissima*, *Nitraria sibirica* and *Reaumuria soongorica* in the southeastern Junggar Basin. *Chinese J. Plant Ecol.* **37**, 665–673 (2013).
- Lin, Z., Xing, X. & Gui-Lian, M. Water sources of shrubs grown in the northern Ningxia Plain of China characterized by shallow groundwater table. *Chinese J. Plant Ecol.* **36**, 618–628 (2012).
- Bijoor, N. S., McCarthy, H. R., Zhang, D. & Pataki, D. E. Water sources of urban trees in the Los Angeles metropolitan area. *Urban Ecosyst.* **15**, 195–214 (2012).
- February, E. C., West, A. G. & Newton, R. J. The relationship between rainfall, water source and growth for an endangered tree. *Austral Ecol.* **32**, 397–402 (2007).

55. Kurz-Besson, C. *et al.* Hydraulic lift in cork oak trees in a savannah-type Mediterranean ecosystem and its contribution to the local water balance. *Plant Soil* **282**, 361–378 (2006).
56. Swaffer, B. A., Holland, K. L., Doody, T. M., Li, C. & Hutson, J. Water use strategies of two co-occurring tree species in a semi-arid karst environment. *Hydrol. Processes* **28**, 2003–2017 (2014).
57. West, A. G. *et al.* Diverse functional responses to drought in a Mediterranean-type shrubland in South Africa. *New Phytol.* **195**, 396–407 (2012).
58. Ohte, N. *et al.* Water utilization of natural and planted trees in the semiarid desert of Inner Mongolia, China. *Ecol. Appl.* **13**, 337–351 (2003).
59. Sun, S., Huang, J., Han, X. & Lin, G. Comparisons in water relations of plants between newly formed riparian and non-riparian habitats along the bank of Three Gorges Reservoir, China. *Trees Struct. Funct.* **22**, 717–728 (2008).
60. Berry, Z. C., Hughes, N. M. & Smith, W. K. Cloud immersion: an important water source for spruce and fir saplings in the southern Appalachian Mountains. *Oecologia* **174**, 319–326 (2014).
61. Jia, G., Yu, X., Deng, W., Liu, Y. & Li, Y. Determination of minimum extraction times for water of plants and soils used in isotopic analysis. *J. Food Agric. Environ.* **10**, 1035–1040 (2012).
62. Rong, L., Chen, X., Chen, X., Wang, S. & Du, X. Isotopic analysis of water sources of mountainous plant uptake in a karst plateau of southwest China. *Hydrol. Processes* **25**, 3666–3675 (2011).
63. Tang, K. L. & Feng, X. H. The effect of soil hydrology on the oxygen and hydrogen isotopic compositions of plants' source water. *Earth Planet. Sci. Lett.* **185**, 355–367 (2001).
64. Wang, P., Song, X., Han, D., Zhang, Y. & Liu, X. A study of root water uptake of crops indicated by hydrogen and oxygen stable isotopes: a case in Shanxi Province, China. *Agric. Water Manage.* **97**, 475–482 (2010).
65. Wei, Y. F., Fang, J., Liu, S., Zhao, X. Y. & Li, S. G. Stable isotopic observation of water use sources of *Pinus sylvestris* var. *mongolica* in Horqin Sandy Land, China. *Trees Struct. Funct.* **27**, 1249–1260 (2013).
66. Wei, L., Lockington, D. A., Poh, S., Gasparon, M. & Lovelock, C. E. Water use patterns of estuarine vegetation in a tidal creek system. *Oecologia* **172**, 485–494 (2013).
67. Zhang, W. *et al.* Using stable isotopes to determine the water sources in alpine ecosystems on the east Qinghai-Tibet plateau, China. *Hydrol. Processes* **24**, 3270–3280 (2010).
68. Anderegg, L. D. L., Anderegg, W. R. L., Abatzoglou, J., Hausladen, A. M. & Berry, J. A. Drought characteristics' role in widespread aspen forest mortality across Colorado, USA. *Glob. Change Biol.* **19**, 1526–1537 (2013).
69. Berkelhammer, M. *et al.* The nocturnal water cycle in an open-canopy forest. *J. Geophys. Res.* **D 118**, 10225–10242 (2013).
70. Bertrand, G. *et al.* Determination of spatiotemporal variability of tree water uptake using stable isotopes ($\delta^{18}\text{O}$, $\delta^2\text{H}$) in an alluvial system supplied by a high-altitude watershed, Pfyn forest, Switzerland. *Ecohydrol.* **7**, 319–333 (2014).
71. Liu, Y. *et al.* Analyzing relationships among water uptake patterns, rootlet biomass distribution and soil water content profile in a subalpine shrubland using water isotopes. *Eur. J. Soil Biol.* **47**, 380–386 (2011).
72. Penna, D. *et al.* Tracing the water sources of trees and streams: isotopic analysis in a small pre-alpine catchment. *Proc. Env. Sci.* **19**, 106–112 (2013).
73. Phillips, S. L. & Ehleringer, J. R. Limited uptake of summer precipitation by bigtooth maple (*Acer grandidentatum* Nutt) and Gambel's Oak (*Quercus gambelii* Nutt). *Trees Struct. Funct.* **9**, 214–219 (1995).
74. Rose, K. L., Graham, R. C. & Parker, D. R. Water source utilization by *Pinus jeffreyi* and *Arctostaphylos patula* on thin soils over bedrock. *Oecologia* **134**, 46–54 (2003).
75. Brunel, J. P., Walker, G. R. & Kennett-Smith, A. K. Field validation of isotopic procedures for determining sources of water used by plants in a semiarid environment. *J. Hydrol.* **167**, 351–368 (1995).
76. Clinton, B. D., Vose, J. M., Vroblesky, D. A. & Harvey, G. J. Determination of the relative uptake of ground vs. surface water by *Populus deltoides* during phytoremediation. *Int. J. Phytoremed.* **6**, 239–252 (2004).
77. Eggemeyer, K. D. *et al.* Seasonal changes in depth of water uptake for encroaching trees *Juniperus virginiana* and *Pinus ponderosa* and two dominant C(4) grasses in a semiarid grassland. *Tree Physiol.* **29**, 157–169 (2009).
78. Holland, K. L., Tyerman, S. D., Mensforth, L. J. & Walker, G. R. Tree water sources over shallow, saline groundwater in the lower River Murray, south-eastern Australia: implications for groundwater recharge mechanisms. *Aust. J. Bot.* **54**, 193–205 (2006).
79. Kukowski, K. R., Schwinning, S. & Schwartz, B. F. Hydraulic responses to extreme drought conditions in three co-dominant tree species in shallow soil over bedrock. *Oecologia* **171**, 819–830 (2013).
80. McCole, A. A. & Stern, L. A. Seasonal water use patterns of *Juniperus ashei* on the Edwards Plateau, Texas, based on stable isotopes in water. *J. Hydrol.* **342**, 238–248 (2007).
81. Mensforth, L. J., Thorburn, P. J., Tyerman, S. D. & Walker, G. R. Sources of water used by riparian Eucalyptus-Camaldulensis overlying highly saline groundwater. *Oecologia* **100**, 21–28 (1994).
82. Hartsough, P., Poulson, S. R., Biondi, F. & Estrada, I. G. Stable isotope characterization of the ecohydrological cycle at a tropical treeline site. *Arct. Antarct. Alp. Res.* **40**, 343–354 (2008).
83. Brunel, J. P., Walker, G. R., Dighton, J. C. & Monteny, B. Use of stable isotopes of water to determine the origin of water used by the vegetation and to partition evapotranspiration. A case study from HAPEX-Sahel. *J. Hydrol. (Amst.)* **188–189**, 466–481 (1997).
84. February, E. C., Higgins, S. I., Newton, R. & West, A. G. Tree distribution on a steep environmental gradient in an arid savanna. *J. Biogeogr.* **34**, 270–278 (2007).
85. Garcin, Y. *et al.* Hydrogen isotope ratios of lacustrine sedimentary n-alkanes as proxies of tropical African hydrology: insights from a calibration transect across Cameroon. *Geochim. Cosmochim. Acta* **79**, 106–126 (2012).
86. Deng, Y., Jiang, Z. & Qin, X. Water source partitioning among trees growing on carbonate rock in a subtropical region of Guangxi, China. *Env. Earth Sci.* **66**, 635–640 (2012).
87. Evaristo, J. A., McDonnell, J. J. & Scholl, M. A. Evidence for ecohydrological separation across contrasting sites in a wet tropical low seasonality catchment. *Hydrol. Processes* (submitted).
88. Nie, Y. *et al.* Seasonal water use patterns of woody species growing on the continuous dolostone outcrops and nearby thin soils in subtropical China. *Plant Soil* **341**, 399–412 (2011).
89. Rosado, B. H. P., De Mattos, E. A. & Sternberg, L. D. S. L. Are leaf physiological traits related to leaf water isotopic enrichment in restinga woody species? *An. Acad. Bras. Cienc.* **85**, 1035–1045 (2013).
90. Schwendenmann, L., Pendall, E., Sanchez-Bragado, R., Kunert, N. & Holsher, D. Tree water uptake in a tropical plantation varying in tree diversity: interspecific differences, seasonal shifts and complementarity. *Ecohydrol.* **8**, 1–12 (2014).

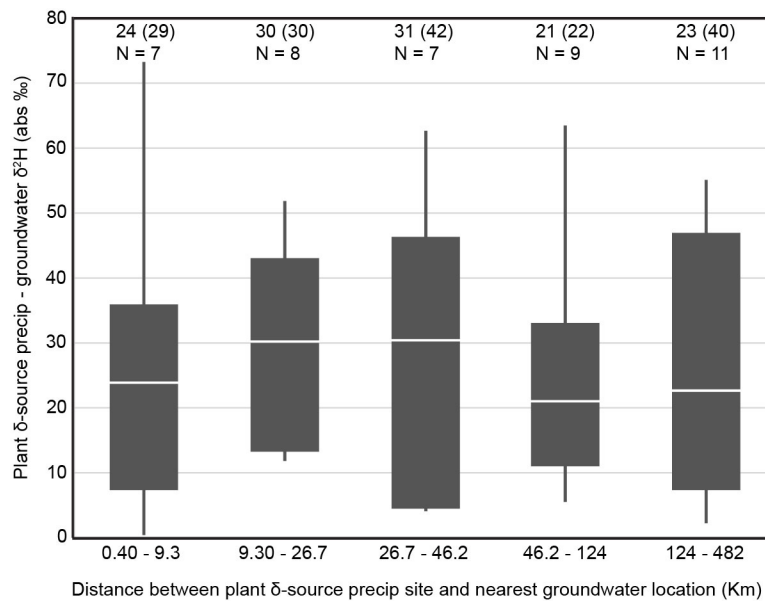


Extended Data Figure 1 | Schematic representation of tracing the isotopic composition of source precipitation. Plant xylem water isotopic values plot on a linear regression called the evaporation line. The point on the local meteoric water line (LMWL) where the plant xylem water evaporation line intersects provide a good approximation of the mean isotopic value of plant xylem source precipitation. The same method is used in tracing the soil water δ source value.



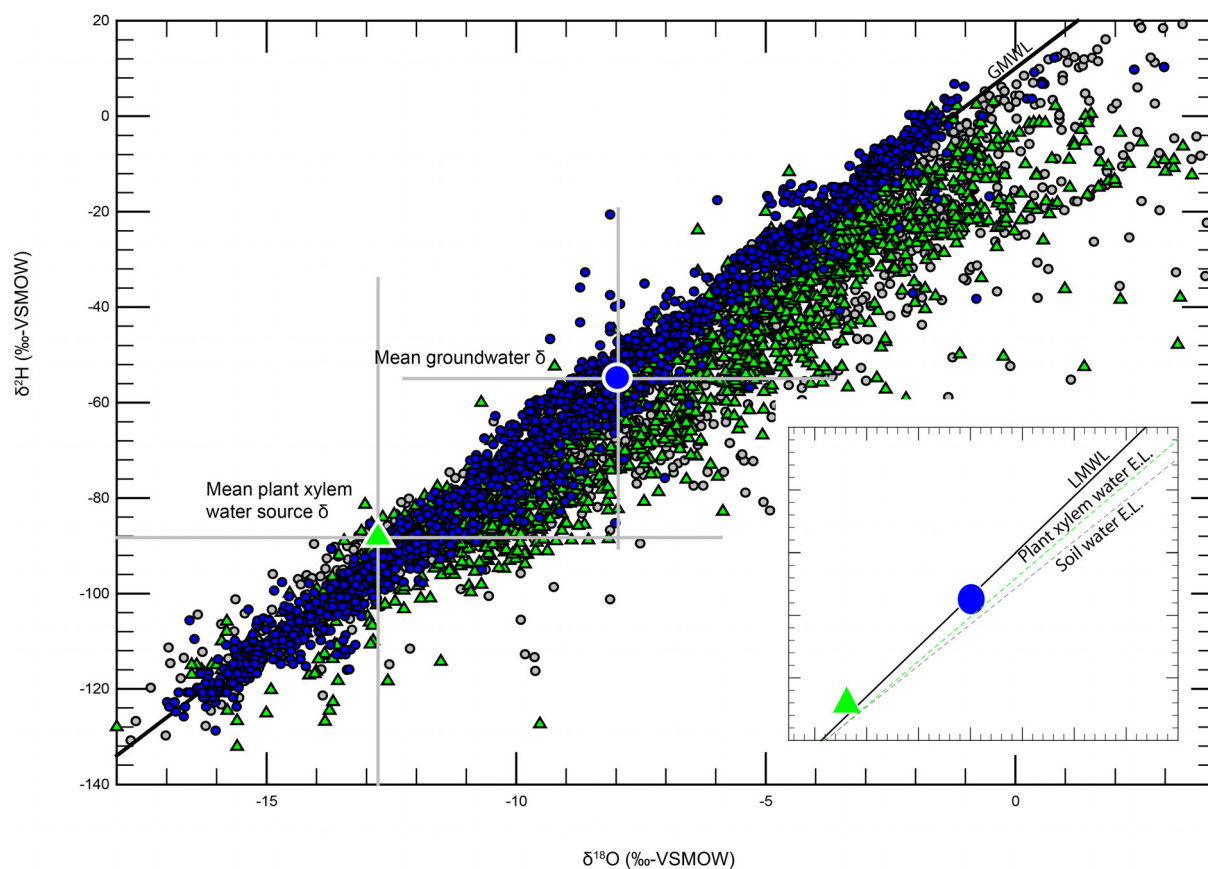
Extended Data Figure 2 | Tracing the isotopic composition of plant xylem source precipitation versus mean groundwater value. Plant xylem water (grey triangles, $n = 88$) plotted in $\delta^{18}\text{O}$ – $\delta^2\text{H}$ space. Shown are the mean plant xylem source precipitation value (green triangle with error bars, ± 1 s.d., $n = 88$), mean groundwater value (blue circle with error bars, ± 1 s.d., $n = 271$),

amount-weighted average precipitation (yellow star), GMWL (solid black line) and LMWL (dashed black line). This is an example of a case in Oregon, USA (ref. 1) where mean groundwater isotope value is more positive than plant xylem source precipitation value. This is the case in 41 of 47 sites in our database.



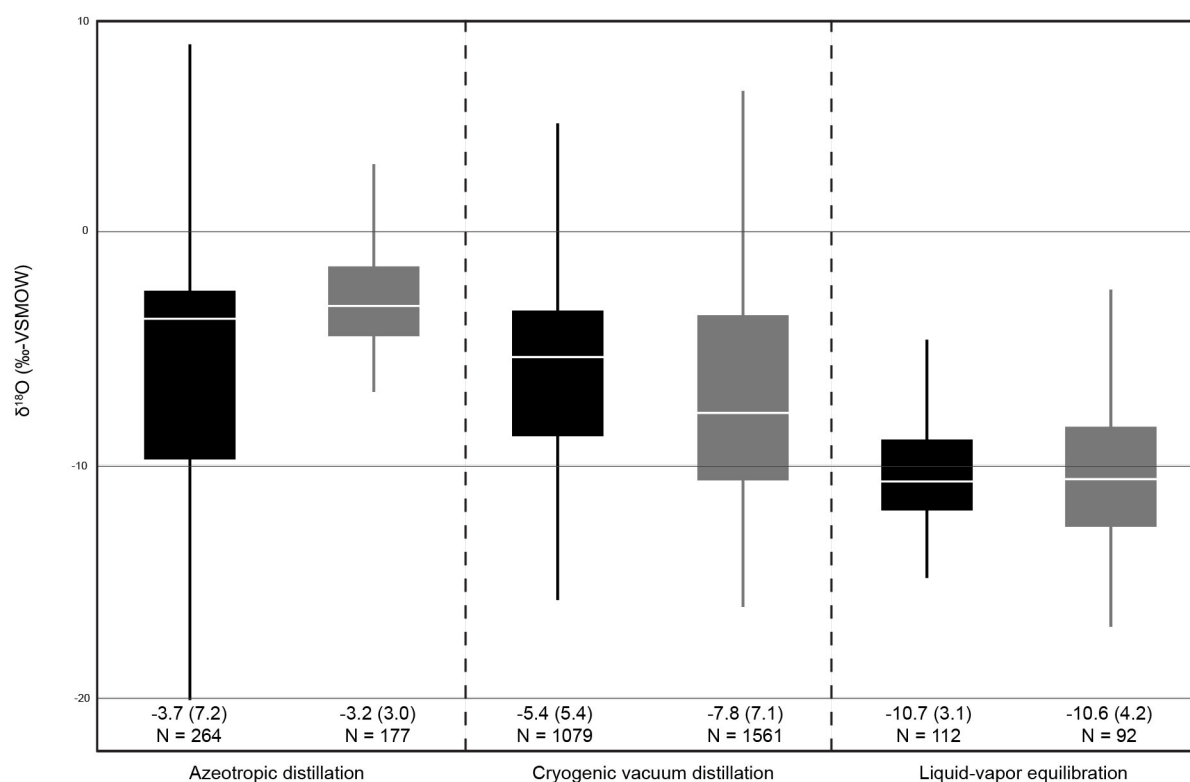
Extended Data Figure 3 | The difference between plant xylem δ -source precipitation values and mean groundwater $\delta^2\text{H}$ values, plotted against increasing distance of groundwater locations from actual plant xylem study sites. The extents of the boxes show the 25th and 75th percentiles; whiskers

show the extents of outliers. Also shown are median (interquartile range) values ($P > 0.90$, Tukey–Kramer honest significant difference) for five ($n = 7$; $n = 8$; $n = 7$; $n = 9$; $n = 11$) arbitrary distance ranges.



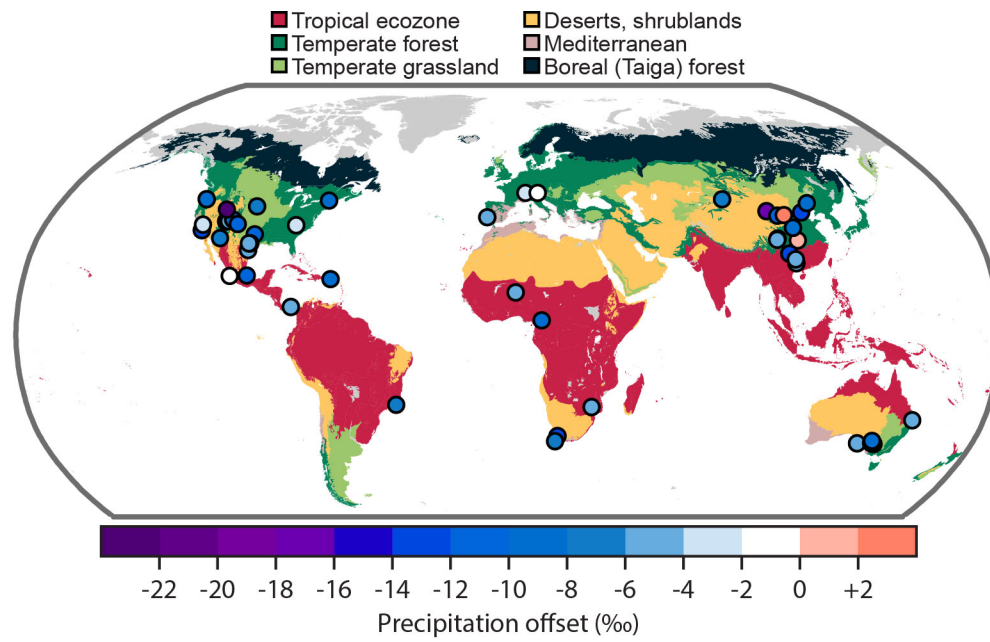
Extended Data Figure 4 | Groundwater and plant xylem source precipitation. Plot of $\delta^{18}\text{O}$ versus $\delta^2\text{H}$ for global plant xylem water (green triangles, $n = 1,460$), soil water (grey circles, $n = 1,830$), and groundwater (blue circles, $n = 2,749$). Also shown are the isotopic composition of source precipitation that leads to groundwater recharge (blue circle with error bars,

mean ± 1 s.d.) and precipitation that leads to plant water uptake (green triangle with error bars, mean ± 1 s.d.). The inset shows the linear regression of plant xylem water and soil water, forming distinct evaporation lines (ELs) whereby, at a site level, plant xylem water is completely bounded by soil water. Also shown are GMWL and LMWL in the main plot and inset, respectively.



Extended Data Figure 5 | Comparison of plant xylem (black boxes) and soil water (grey boxes) $\delta^{18}\text{O}$, based on water extraction techniques. Cryogenic vacuum ($n = 2,640$) and azeotropic distillation ($n = 441$) are significantly different from liquid-vapour equilibration methods ($n = 204$) ($P < 0.0001$, nonparametric Dunn method for joint ranking). Cryogenic vacuum and

azeotropic distillation are not significantly different from each other ($P = 0.35$, nonparametric Dunn method for joint ranking). The extents of the boxes show the 25th and 75th percentiles; whiskers show the extents of outliers. Also shown are median (interquartile range) values for each water type and water extraction technique.



Extended Data Figure 6 | Global map of plant xylem water precipitation offsets from 47 study sites.

Extended Data Table 1 | Site-by-site source precipitation δ values for plant xylem water, groundwater and soil water

Site ID	Location	Plant xylem δ source ($\delta^2\text{H}$)	Groundwater ($\delta^2\text{H}$)	Soil water δ source ($\delta^2\text{H}$)	Plant vs. Groundwater	Soil vs. Groundwater
1	Texas A, US ⁴⁶	-38 (27)	-25 (1.8)	-33 (27)	N.S.	**
2	Arizona A, US ⁴⁷	-131 (16)	-96 (22)	-91 (11)	**	N.S.
3	Arizona B, US ⁴⁸	-96 (43)	-56 (13)	-45 (53)	**	N.S.
4	Arizona-Utah, US ⁴⁹	-125 (30)	-81 (0.5)	-160 (24)	**	**
5	Northwest CN ⁵⁰	-91 (54)	-81 (17)	-30 (49)	N.S.	N.S.
6	Junggar Plain, CN ⁵¹	-109 (9.9)	-49 (13)	-87 (40)	**	**
7	Ningxia Plain, CN ⁵²	-139 (26)	-76 (15)	-149 (21)	**	**
8	California A, US ⁵³	-86 (31)	-44 (5.5)	-57 (43)	**	**
9	Cape Town, ZA ⁵⁴	-58 (29)	-33 (20)	-	**	-
10	Evora, PT ⁵⁵	-46 (2.7)	-29 (2.3)	-51 (10)	**	**
11	Eyre Peninsula, AU ⁵⁶	-31 (10)	-24 (1)	-15 (6.4)	**	**
12	Mt. Natl Park, ZA ⁵⁷	-18 (13)	-18 (16)	-	N.S.	-
13	Ordos Plateau, CN ⁵⁸	-113 (33)	-63 (13)	-94 (34)	**	**
14	Hubei Province, CN ⁵⁹	-114 (34)	-66 (6.8)	-125 (50)	**	**
15	N. Carolina, US ⁶⁰	-52 (13)	-40 (4.1)	-56 (12)	**	**
16	Beijing, CN ⁶¹	-90 (6.4)	-64 (23)	-102 (8.1)	**	**
17	Guizhou Prov, CN ⁶²	-152 (51)	-49 (7.2)	-145 (54)	**	**
18	New Hampshire, US ⁶³	-103 (29)	-52 (9.2)	-	**	-
19	Shanxi Prov, CN ⁶⁴	-98 (8.8)	-62 (4.8)	-108 (29)	**	**
20	Horqin, CN ⁶⁵	-147 (21)	-75 (8)	-141 (41)	**	**
21	Queensland, AU ⁶⁶	-18 (23)	-30 (3.3)	-	N.S.	-
22	Sichuan A, CN ⁶⁷	-120 (9.6)	-86 (12)	-131 (33)	**	**
23	Colorado A, US ⁶⁸	-214 (34)	-107 (17)	-190 (38)	**	**
24	Colorado B, US ⁶⁹	-156 (141)	-92 (13)	-	**	-
25	Sierre, CH ⁷⁰	-151 (39)	-106 (5.8)	-142 (56)	**	**
26	Oregon, US ¹	-130 (10)	-100 (53)	-132 (36)	**	**
27	Sichuan B, CN ⁷¹	-127 (10)	-86 (12)	-124 (11)	**	**
28	Pre-Alpine, IT ⁷²	-	-56 (2.1)	-	-	-
29	Utah, US ⁷³	-128 (69)	-114 (23)	-	N.S.	-
30	California B, US ⁷⁴	-128 (21)	-92 (8.4)	-126 (49)	**	**
31	Victoria, AU ⁷⁵	-29 (32)	-35 (0.3)	-43 (22)	N.S.	N.S.
32	Texas B, US ⁷⁶	-64 (12)	-29 (2.6)	-	**	-
33	Nebraska, US ⁷⁷	-134 (42)	-73 (11)	-	N.S.	-
34	River Murray A, AU ⁷⁸	-14 (40)	-33 (7)	-29 (49)	N.S.	N.S.
35	Texas C, US ⁷⁹	-25 (18)	-17 (1.6)	-	**	-
36	Texas D, US ⁸⁰	-66 (37)	-26 (3.8)	-40 (29)	**	**
37	River Murray B, AU ⁸¹	-42 (7.7)	-30 (3.3)	-36 (14)	**	**
38	West Central, MX ⁸²	-216 (43)	-71 (23)	-242 (105)	**	**
39	Niger ⁸³	-28 (4.9)	-41 (16)	-16 (25)	**	**
40	Limpopo, ZA ⁸⁴	-64 (19)	-19 (15)	-	**	-
41	Cameroon ⁸⁵	-45 (30)	-15 (2.7)	-	**	-
42	Guangxi A, CN ⁸⁶	-146 (26)	-49 (14)	-104 (119)	**	N.S.
43	Luquillo-Susua, PR ⁸⁷	-27 (19)	-8.8 (7.5)	-14 (26)	**	**
44	Veracruz, MX ²	-67 (23)	-71 (23)	-126 (53)	N.S.	**
45	Guangxi B, CN ⁸⁸	-105 (45)	-49 (7.2)	-142 (113)	**	**
46	Rio de Janeiro, BR ⁸⁹	-32 (44)	-15 (2.7)	10.4 (23)	**	**
47	Sardinilla, PA ⁹⁰	-54 (22)	-15 (2.7)	-58 (29)	**	**

Plant xylem and soil water δ source precipitation values (median, interquartile range) are calculated using equations (2) and (3). The last two columns show whether or not the source precipitation values are statistically different amongst the three water compartments. N.S., not significant. Superscript numbers after site locations refer to the source paper (refs 46–90).

**Denotes statistically significant difference ($\alpha = 0.05$).

Extended Data Table 2 | Site-by-site soil water precipitation offset values

Site ID	Precip offset _{soil}
1	-6.2 (1.1)
2	-0.2 (1.8)
3	-4.1 (1.6)
4	-7.4 (1.9)
5	-0.7 (4)
6	-5.6 (5.4)
7	-1.5 (0.6)
8	-7.5 (1.6)
10	-5.5 (3)
11	-3.6 (0.2)
13	-0.8 (1.2)
14	-4.5 (1)
15	-3 (0.9)
16	-1.4 (0.7)
17	-6 (1.1)
19	-4.5 (1.3)
20	1.78 (1.2)
22	-9.7 (1.2)
23	-6.5 (0.6)
25	-1.1 (1.1)
26	-5.6 (1)
27	-4.7 (0.6)
28	-6.4 (0.3)
30	-4 (2.4)
31	-5.4 (1)
34	-5.5 (1)
36	-4.1 (1.1)
37	-5 (1.3)
38	-3.9 (0.8)
39	-2.4 (2.5)
42	-12 (3.3)
43	-8.2 (0.9)
44	-10 (1.2)
45	-8.4 (1.4)
46	-9.9 (2.5)
47	-3.3 (0.8)

Values are median (interquartile range).

Extended Data Table 3 | Biome-level soil water precipitation offset values

Biome	Precip offset _{soil}
Arid	-3.7 (4.6)
Mediterranean	-7 (3.6)
Temperate forests	-5.2 (2.4)
Temperate grasslands	-5.1 (1.1)
Tropics	-9.3 (2.2)

Values are median (interquartile range).

Broad plumes rooted at the base of the Earth's mantle beneath major hotspots

Scott W. French^{1†} & Barbara Romanowicz^{1,2,3}

Plumes of hot upwelling rock rooted in the deep mantle have been proposed as a possible origin of hotspot volcanoes, but this idea is the subject of vigorous debate^{1,2}. On the basis of geodynamic computations, plumes of purely thermal origin should comprise thin tails, only several hundred kilometres wide³, and be difficult to detect using standard seismic tomography techniques. Here we describe the use of a whole-mantle seismic imaging technique—combining accurate wavefield computations with information contained in whole seismic waveforms⁴—that reveals the presence of broad (not thin), quasi-vertical conduits beneath many prominent hotspots. These conduits extend from the core–mantle boundary to about 1,000 kilometres below Earth's surface, where some are deflected horizontally, as though entrained into more vigorous upper-mantle circulation. At the base of the mantle, these conduits are rooted in patches of greatly reduced shear velocity that, in the case of Hawaii, Iceland and Samoa, correspond to the locations of known large ultralow-velocity zones^{5–7}. This correspondence clearly establishes a continuous connection between such zones and mantle plumes. We also show that the imaged conduits are robustly broader than classical thermal plume tails, suggesting that they are long-lived⁸, and may have a thermochemical origin^{9–11}. Their vertical orientation suggests very sluggish background circulation below depths of 1,000 kilometres. Our results should provide constraints on studies of viscosity layering of Earth's mantle and guide further research into thermochemical convection.

More than 40 years ago, Morgan¹ proposed that hotspot volcanoes are the surface expression of narrow plumes of hot material that originate in a boundary layer in the deep mantle, as one would expect of a convecting fluid that is heated from below³. Whether deep mantle plumes exist, and how deep their roots are, has been the subject of lively debate, which continues to this day. The idea that hotspots may be anchored at the core–mantle boundary (CMB) is supported by several observations: the relative fixity of hotspots with respect to global mantle circulation¹; the correlation of hotspot locations with the large low shear velocity provinces (LLSVPs) at the base of the mantle¹²; and a suggestion from geodynamic modelling¹³ that hotspots might preferentially occur above ultralow velocity zones (ULVZs). A radically different origin for hotspots has also been proposed, in which these features are the consequence of melting owing to shallow convective processes, with their morphologies controlled by stresses and cracks within the lithosphere².

In the classical view³, a mantle plume is composed of a large head and a thin tail, which connects it to a root deeper in the mantle. If such plumes were to originate at the base of the mantle, we would expect the lower mantle to contain narrow (less than 200 km in diameter, according to relevant scaling relations¹⁴), continuous, vertically oriented columns of hotter-than-average (and therefore of low seismic velocity) material, located in the vicinity of presently active hotspots.

In the deep mantle, short-wavelength, low-velocity anomalies are difficult to image with standard seismic tomographic techniques, which typically rely on travel times of body waves (seismic waves that travel

through the interior of Earth); such anomalies can be hidden from view by wavefront healing effects¹⁵. Also, most hotspot volcanoes, and potentially any associated plumes, are located in the middle of oceans, where they are difficult to image owing to the lack both of dense seismic networks and of earthquakes with the appropriate geometry.

Thus, while various tomographic studies have hinted at the presence of plume-like features in the lower mantle associated with some subset of the major hotspots^{16–18}, ambiguity remains as to the vertical continuity of these features, how distinct they are from other low-velocity 'blobs' in the lower mantle, and whether they represent detection of the narrow type of plumes typically associated with purely thermal convection¹⁹. To improve the resolution of low-velocity features of limited lateral extent such as plumes, two ingredients are needed: first, better illumination of Earth's interior; and second, improved theoretical description of the interaction of the seismic wavefield with the three-dimensional Earth structure.

Here we present robust evidence for large, vertically continuous, low-velocity columns in the lower mantle beneath many prominent hotspots, from our recent global, radially anisotropic, whole-mantle shear-wave velocity model, SEMUCB-WM1 (ref. 4). This model was constructed by inversion of a large data set of full, long-period seismograms, including first- and second-orbit fundamental mode and overtone surface waves down to 60 s, as well as body waveforms down to 32 s. Because it includes surface-wave overtones, shear waves diffracted along the core–mantle boundary (S_{diff}) and multiply-reflected waves between the surface and the CMB, this data set provides considerably better illumination of the whole-mantle volume than can be obtained with a standard set of travel times alone. In addition, accurate numerical computation of the forward wavefield using the spectral element method²⁰ at each iteration of the model construction allows us to better resolve regions of lower-than-average shear-wave velocity, as previously illustrated for the upper mantle²¹. The construction of this model is briefly summarized in the Methods.

In model SEMUCB-WM1 (ref. 4), broad, dome-like plumes that show a reduction of shear-wave velocity by more than 1.5%–2% are present in the lower mantle beneath Samoa, Hawaii and the Pacific Superswell volcanoes. These plumes are clearly distinct from other, more isolated and weaker low-velocity features that appear in cross-sections spanning half of Earth's circumference (Fig. 1). The plumes are rooted in patches of more strongly reduced shear-wave velocity near the CMB, and extend vertically up to depths of at least 1,000 km, above which their character changes. A three-dimensional view of the central Pacific region (Fig. 2) shows that the cores of these plumes are well separated from each other across most of the lower mantle, embedded in the lower-than-average-velocity background of the Pacific LLSVP. This is particularly clear in the depth range 1,000 km to 1,500 km, where there is a one-to-one relationship with the corresponding hotspot volcanoes, although the plumes are not always located exactly beneath the volcanoes. Comparison with previous global models (Extended Data Fig. 1) indicates general agreement on the background long-wavelength features, while in SEMUCB-WM1 the

¹Department of Earth and Planetary Science, Berkeley Seismological Laboratory, University of California at Berkeley, California 94720, USA. ²Institut de Physique du Globe, Paris 75238, France. ³Collège de France, Paris 75005, France. [†]Present address: National Energy Research Scientific Computing Center, Lawrence Berkeley National Laboratory, Berkeley, California 94720, USA.

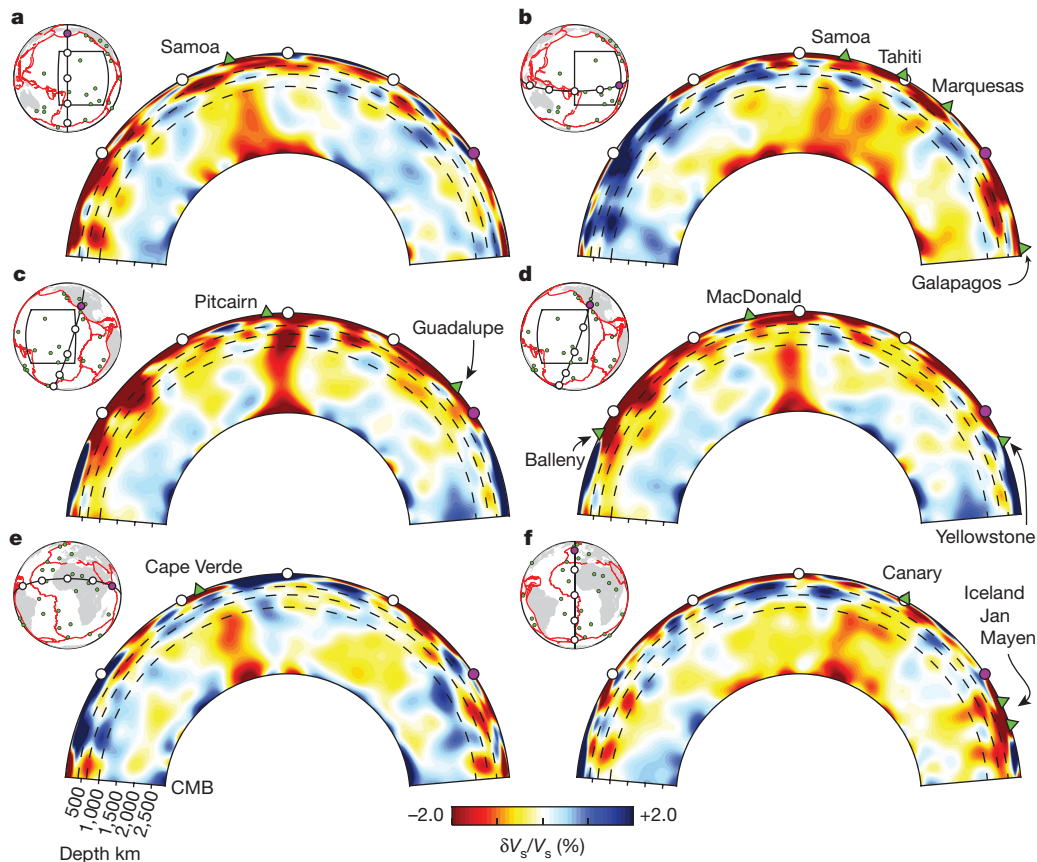


Figure 1 | Whole-mantle depth cross-sections of relative shear-velocity variations in model SEMUCB-WM1⁴, in the vicinity of major hotspots. The sections are shown in the inset maps, with the direction of the projection indicated by the position of the purple dot in both map and cross-section views (black boxes correspond to the three-dimensional rendering regions in Fig. 2). Green dots and triangles mark the locations of hotspots²⁷. The reference model is the corresponding global one-dimensional average shear-wave velocity (V_s) profile of SEMUCB-WM1. The colour scale has been chosen to emphasize lower-mantle structures, resulting in substantial saturation in the upper mantle. Broken lines indicate depths of 410 km, 660 km and 1,000 km. Focused, quasi-vertical, broad plumes extend continuously from

plumes stand out as continuous features confined to well defined vertically oriented columns.

In particular, the Hawaiian plume appears as a separate vertical conduit of varying width (Fig. 2a–c), with a weaker zone at around 500 km above the CMB, rooted in its own patch of strongly reduced shear velocity at the base of the mantle. In the transition zone, this plume appears to be strongly deflected towards the west–southwest (Fig. 3). This morphology is compatible with evidence for a hot upper mantle to the west of Hawaii, based on the analysis of converted waves (that is, receiver functions)²². The presence of bodies with higher-than-average velocity southwest and northeast of the Hawaiian chain is in agreement with regional studies^{23,24}. However, in the lower mantle, the associated conduit is more vertically oriented in SEMUCB-WM1. Similar broad, vertically oriented low-velocity conduits are found in the vicinity of some hotspots lying on the border of the African LLSVP (Figs 1e, f and 3d, e and Extended Data Fig. 2).

The lower-mantle plume conduits described above are rooted in wide patches (of diameter 500–800 km) of strongly negative velocity reduction near the CMB. In at least three cases, these patches coincide in location with large ULVZs previously detected in the vicinity of the corresponding hotspots: near Hawaii (detected through observations of post-cursors to diffracted S waves⁵), and beneath Iceland⁶ and Samoa⁷ (found through the study of waveform distortion in the phase SP_{diffKS}).

patches of strongly reduced V_s at the base of the mantle to depths of at least 1,000 km in the vicinity of: **a**, Samoa; **b**, Tahiti, the Marquesas, the Galapagos and Samoa; **c**, Pitcairn; **d**, MacDonal; **e**, Cape Verde; and **f**, the Canary Islands. These plumes stand out from other low-velocity features in these cross-sections, which span nearly half of Earth. **d**, Note the absence of a noticeable anomaly in the lower mantle immediately beneath the Yellowstone hotspot. However, a faint low-velocity conduit appears to the southwest (offshore of North America), anchored by a low-velocity patch in the D'' mantle region. It is beyond the resolution of our study to verify whether this feature is related to the Yellowstone or the Guadalupe (c) hotspot.

Because of the computational challenges of whole-mantle imaging using full waveforms and numerical simulations, the resolution of our model is limited by our choice of parametrization and maximum frequency. However, resolution tests (Methods; Extended Data Figs 4–8) clearly indicate that our approach can resolve the vertical continuity of plumes without ray-like smearing or erroneous deflection, and that the variations of the shape and amplitude of the plumes with depth are likely to be robust features. These tests also indicate (see Supplementary Information section S1 and Supplementary Figs 1 and 2) that our modelling approach can distinguish between hypothetical broad superplume-like features and the distinct vertical conduits that are shown in Fig. 1. Numerical experiments (Supplementary Information section S2 and Supplementary Figs 5–9) demonstrate that plumes of the same scale as are seen in Fig. 1 and used in our tests should be readily detectable in the waveform data used by our inversion. Furthermore, on the basis of relative amplitude recovery alone, our resolution tests also show that in order to obtain a velocity reduction of 2% or more over the major part of the lower mantle—as seen in our model—a narrow plume would have to be very strong (that is, >10% reduction in shear-wave velocity for a plume of width <200 km; see Methods and Extended Data Fig. 4). Such a strong velocity contrast would translate into unrealistically high²⁵ effective temperature excesses of 1,500–2,000 °C. In contrast, for a 2% velocity anomaly over a width of 800–1,000 km, as imaged in SEMUCB-WM1 under Hawaii

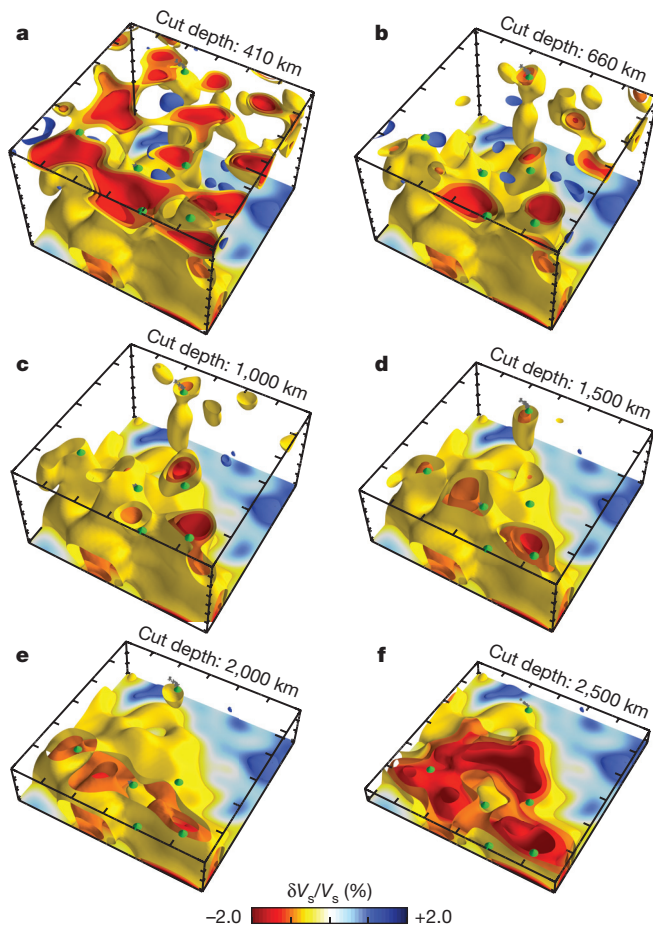


Figure 2 | Three-dimensional rendering of shear-wave-velocity structure in the Pacific Superswell region. Relative velocity perturbations are shown with respect to the global average at each depth. For each panel, the location of the box is shown in the inset map of Fig. 1. The region is shown from above, with cuts at increasing depths: **a**, 410 km; **b**, 660 km; **c**, 1,000 km; **d**, 1,500 km; **e**, 2,000 km; **f**, 2,500 km. The following hotspot locations, projected down from the surface, are indicated by green cones in each box: Hawaii (top; north), Samoa (left; west), and the four Superswell hotspots: Tahiti, Pitcairn, Marquesas and MacDonald. Well defined vertically oriented conduits with central cores of velocity lower than -1.5% can be associated with each of the hotspots, particularly clearly in **c** and **d**. The low-velocity conduit beneath Hawaii stands out in **b–d**. In **f**, patches of much lower-than-average velocity start appearing within the Pacific LLSVP, continuing down to the CMB. On the other hand, at a depth of 410 km, the low-velocity conduits start spreading horizontally and merge into the depth range in which low-velocity fingers have previously been observed in the upper mantle²¹.

(Fig. 3), effective temperature excesses of 400–500 °C (estimated in a similar manner) are plausible, if one considers that the plumes are rooted in a chemically dense layer⁹. Whether or not these plumes entrain much chemical heterogeneity to the upper mantle is beyond the resolution of this study; however, their width and vertically varying shapes are compatible with models of thermochemical plumes^{10,11}.

Clearly resolved plumes with similar characteristics are found beneath 11 major hotspots (Fig. 4 and Extended Data Table 1). Visual inspection of the SEMUCB-WM1 model suggests the presence of weaker conduits extending from the CMB through most of the lower mantle beneath several other hotspots. Taken together, this ensemble of hotspots includes all of those classified as ‘primary’ in ref. 26. Interestingly, all of the plumes we detected are located within or at the borders of the African and Pacific LLSVPs. In contrast, we found no deep-mantle expression of those hotspots^{26,27} that are located above faster-than-average shear-wave velocities in the mantle region next to the CMB (the D'' region). This may indicate either that the

corresponding plumes are below the detection capabilities of our current modelling approach in the lower mantle, or that they originate in the upper mantle or the uppermost lower mantle. This is the case, in particular, for the Yellowstone hotspot (Fig. 1d). In the case of the Bowie/Juan de Fuca hotspots, the low-velocity anomaly can be followed to the base of the upper mantle²⁸, but not below. Likewise, most of the North African hotspots (except Afar and the neighbouring East African Rift system, particularly beneath Tanzania) have no clear lower-mantle expression. Instead, they appear to derive from a very broad lower-mantle dome, much further south, associated with the African LLSVP (Fig. 3d and Extended Data Fig. 3). The difference in morphology of the Pacific (bundle of plumes) and African (broad dome) LLSVPs, which has been suggested previously²⁹, is now very clear.

Our results confirm the presence of broad plume-like conduits in the lower mantle, located in the vicinity of major hotspots^{4,17}. For the first time, to our knowledge, our study establishes their quasi-vertical continuity from about 1,000-km depth down to the CMB, where they are rooted in patches or domes of much lower-than-average shear velocity, at least some of which (in Iceland, Hawaii and Samoa) coincide with the location of known ULVZs. These plumes exhibit other common characteristics. They are remarkably vertical, indicating that they are not strongly affected by a lower-mantle wind²⁷, and may represent the primary upwellings in the lower mantle. They often have a pinched zone between 500 km and 1,000 km above the CMB—a shape that is similar to what may be expected for thermochemical plumes^{10,11}.

Interestingly, the character of the anomaly appears to change at a depth of around 1,000 km: some plumes are shifted horizontally (for example, Pitcairn and St Helena); others become thinner (such as Samoa and Tahiti); and yet others cannot clearly be tracked to shallower depths (for example, Cape Verde), indicating that they may have split into narrower conduits^{4,18} that, in some cases, are below the resolution of our model. In Iceland and Hawaii (Fig. 3), horizontally elongated arms branch out from the plume stem just below 1,000 km, suggesting that the flow may have encountered resistance to direct continuation into the upper mantle. Similarly, we observe what appears to be ponding of low-velocity material beneath 1,000 km (for example at the Canary Islands; Fig. 1f). This is also the depth range in which some slabs appear to stagnate³⁰. Those plumes that we can track to shallower depths appear to meander through the upper mantle and, in many cases, connect to the previously observed low-velocity channels in the depth range 200–400 km (ref. 21; see, for example, Pitcairn in Figs 1 and 2).

The observed characteristics of these seismically imaged plumes should provide improved constraints on our understanding of the viscosity structure and dynamics of the mantle. The direct connection of these plumes with ULVZs on the one hand, and with prominent hotspots on the other, suggests that the plumes are long-lived and contain material originating in the D'' region that is of lower viscosity than the bulk lower mantle⁸. The change of character at a depth of 1,000 km—taken together with the stagnation of some subducted slabs³⁰—suggests the presence of a notable change in viscosity around that depth. The variation in amplitude and width of these velocity anomalies in the lower mantle may be related to the stage of development of the plumes^{10,11}, or to variations of the viscosity structure with depth^{31,32}, or to a combination of viscosity structure and density contrast owing to chemical heterogeneity at the base of the mantle¹¹. The dominantly vertical character of the plumes in the lower mantle indicates that circulation may be very sluggish away from the plumes. The often-invoked mantle wind²⁷ would be largely confined to the upper mantle, where plumes may indeed become entrained with a substantial non-vertical component (for example, at Pitcairn and Hawaii) into the more vigorous secondary-scale convection that is probably driven by plate motions²¹.

Our study demonstrates the potential of combining waveform tomography with accurate modelling of wave propagation in

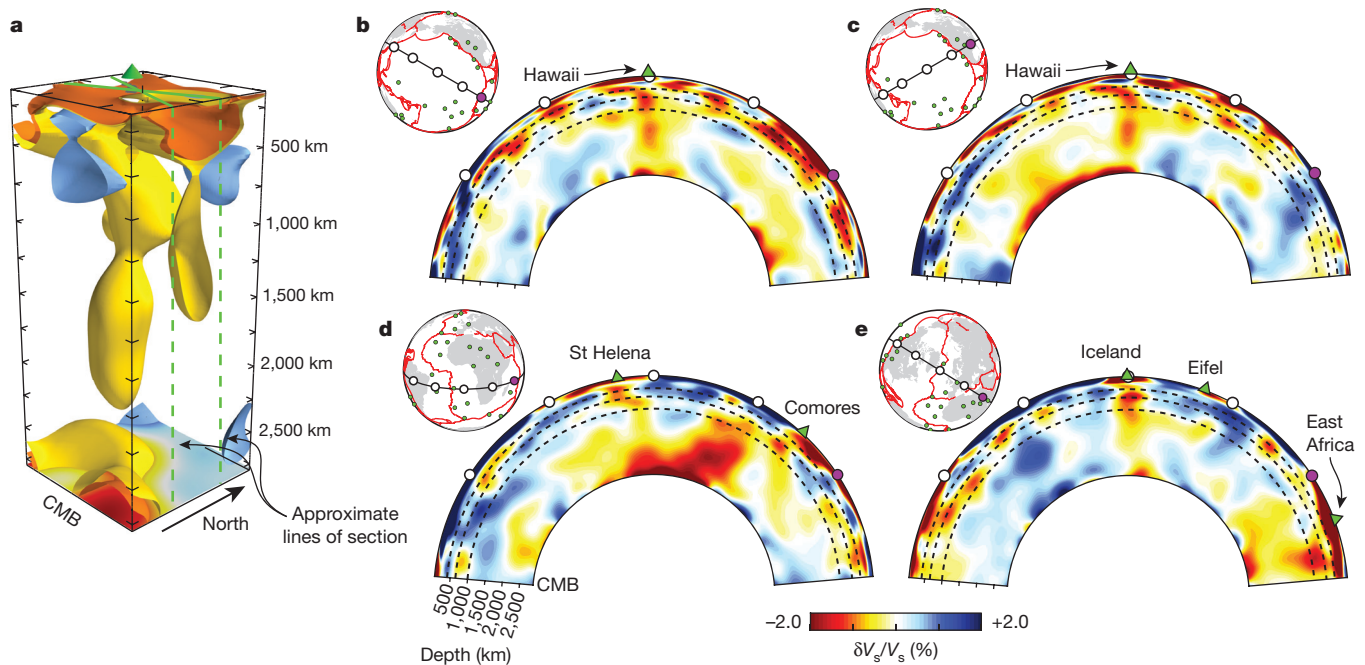


Figure 3 | Hawaii, Iceland, St Helena and the African superplume. **a**, Three-dimensional rendering of the Hawaiian plume, viewed from the southeast. In the lower mantle, the plume conduit is vertically oriented, rooted in a patch of very low V_s at the base of the mantle, with a weak zone in the depth range 2,300–2,600 km. Above approximately 1,000 km, the conduit is deflected towards the west into the transition zone, and appears to interact with the low-velocity finger oriented in the Pacific plate absolute plate motion direction²¹ visible in cross-section above a depth of 500 km. **b–e**, Two-dimensional vertical cross-sections along planes as indicated in the inset maps and corresponding broken green lines in **a**, represented in the same way as in Fig. 1. In **c**, in the

lowermost mantle between the two westernmost reference white dots, we see the edge of the plume associated with the Caroline hotspot. Subduction zones are well imaged in the western Pacific (**b**, **c**), spreading above the 1,000-km horizon, and in south America (**d**); in **c** and **e**, the fossil Farallon subduction extends through the lower mantle. Blue zones in the vicinity of Hawaii in the lower mantle may potentially be downwellings corresponding to other fossil slabs, although this warrants further study. In **e**, we note the particularly broad African plume, one lobe of which extends to the mantle transition zone, the other giving rise in the mid lower mantle to a thinner, weaker plume beneath St Helena.

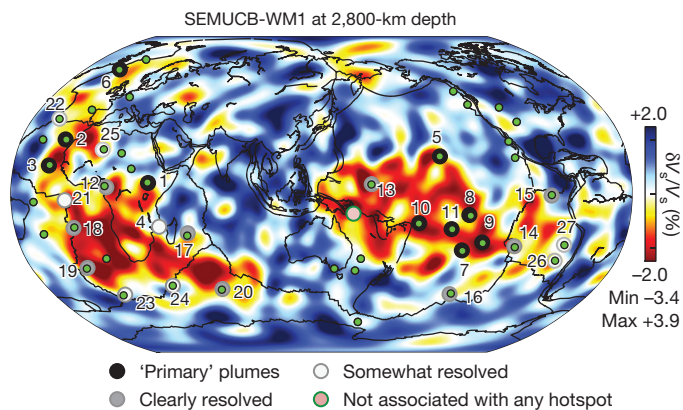


Figure 4 | Locations of plumes detected in the lower mantle in model SEMUCB-WM1⁴. The background map represents the relative V_s variations at 2,800 km in this model, with respect to the global average at that depth. We identify three categories of plumes. ‘Primary’ plumes are those for which $\delta V_s/V_s$ is lower than -1.5% for most of the depth interval 1,000–2,800 km. These 11 plumes also correspond to regions of the lower mantle where the average velocity reduction over the depth range 1,000–1,800 km is significant at the 2σ level (see, for example, Supplementary Figs 3 and 4). Clearly resolved plumes correspond to vertically continuous conduits with $\delta V_s/V_s$ greater than -0.5% in the depth range 1,000–2,800 km. Somewhat resolved plumes have vertically trending conduits with $\delta V_s/V_s$ greater than -0.5% for most of the depth range 1,000–2,800 km, albeit not as clearly continuous. Plumes are numbered as listed in Extended Data Table 1. Green dots represent the global hotspot distribution according to ref. 27. Note that none of the plumes detected falls within a region of faster-than-average velocity at the base of the mantle, and that long-wavelength structure in this model agrees with that of previous tomographic models (see, for example, Supplementary Fig. 10).

order to advance our understanding of the upwelling part of deep mantle flow and its relationship to surface observations. While thin ‘classical’ plume tails may exist in the mantle, they remain below the resolution of global tomography at present. However, our robust confirmation of the existence of broad, possibly thermochemical, plumes—associated with prominent hotspots and rooted in the D’ region (at least some of them in ULVZs)—should provide important constraints for further geodynamical modelling of present-day mantle circulation, and for Earth’s heat-flux budget³³.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 12 November 2014; accepted 19 June 2015.

1. Morgan, W. J. Convection plumes in the lower mantle. *Nature* **230**, 42–43 (1971).
2. Anderson, D. L. Scoring hotspots: the plume and plate paradigms. *Geol. Soc. Special Papers* **388**, 31–54 (2005).
3. Campbell, I. H. & Griffiths, R. W. Implications of mantle plume structure for the evolution of flood basalts. *Earth Planet. Sci. Lett.* **99**, 79–93 (1990).
4. French, S. W. & Romanowicz, B. A. Whole-mantle radially anisotropic shear velocity structure from spectral-element waveform tomography. *Geophys. J. Int.* **199**, 1303–1327 (2014).
5. Cottaar, S. & Romanowicz, B. An unusually large ULVZ at the base of the mantle near Hawaii. *Earth Planet. Sci. Lett.* **355–356**, 213–222 (2012).
6. Helmberger, D. V., Wen, L. & Ding, X. Seismic evidence that the source of the Iceland hotspot lies at the core-mantle boundary. *Nature* **396**, 251–255 (1998).
7. Thorne, M. S., Garnero, E. J., Jahnke, G., Igel, H. & McNamara, A. Mega ultra low velocity zone and mantle flow. *Earth Planet. Sci. Lett.* **364**, 59–67 (2013).
8. Jellinek, A. M. & Manga, M. The influence of a chemical boundary layer on the fixity, spacing and lifetime of mantle plumes. *Nature* **418**, 760–763 (2002).
9. Farnetani, C. G. Excess temperature of mantle plumes: the role of chemical stratification across D’. *Geophys. Res. Lett.* **24**, 1583–1586 (1997).

10. Lin, S.-C. & van Keken, P. E. Dynamics of thermochemical plumes: 2. Complexity of plume structures and its implications for mapping mantle plumes. *Geochim. Geophys. Geosyst.* **7**, Q03003 (2006).
11. Kumagai, I., Davaille, A., Kurita, K. & Stutzmann, E. Mantle plumes: thin, fat, successful or failing? Constraints to explain hot spot volcanism through time and space. *Geophys. Res. Lett.* **35**, L16301 (2008).
12. Thorne, M. S., Garnero, E. J. & Grand, S. P. Geographic correlation between hot spots and deep mantle lateral shear-wave velocity gradients. *Phys. Earth Planet. Inter.* **146**, 47–63 (2004).
13. McNamara, A. K., Garnero, E. J. & Rost, S. Tracking deep mantle reservoirs with ultra-low velocity zones. *Earth Planet. Sci. Lett.* **299**, 1–9 (2010).
14. Davaille, A. & Limare, A. in *Treatise on Geophysics* Vol. 7 (ed. Bercovici, D.) 89–156 (Elsevier, 2007).
15. Nolet, G. & Dahlen, F. A. Wavefront healing and the evolution of seismic delay times. *J. Geophys. Res.* **105**, 19043–19054 (2000).
16. Zhao, D. Global tomographic images of mantle plumes and subducting slabs: insight into deep earth dynamics. *Phys. Earth Planet. Inter.* **146**, 3–34 (2004).
17. Montelli, R. *et al.* Finite-frequency tomography reveals a variety of plumes in the mantle. *Science* **303**, 338–343 (2004).
18. Suetsugu, D. *et al.* South Pacific mantle plumes imaged by seismic observation on islands and seafloor. *Geochim. Geophys. Geosyst.* **10**, Q11014 (2009).
19. Boschi, L., Becker, T. W. & Steinberger, B. On the statistical significance of correlations between synthetic mantle plumes and tomographic models. *Phys. Earth Planet. Inter.* **167**, 230–238 (2008).
20. Capdeville, Y., Chaljub, E., Vilotte, J. P. & Montagner, J. P. Coupling the spectral element method with a modal solution for elastic wave propagation in global earth models. *Geophys. J. Int.* **152**, 34–67 (2003).
21. French, S. W., Lekic, V. & Romanowicz, B. Waveform tomography reveals channelled flow at the base of the oceanic asthenosphere. *Science* **342**, 227–230 (2013).
22. Li, X. *et al.* Mapping the Hawaiian plume conduit with converted seismic waves. *Nature* **405**, 939–941 (2000).
23. Wolfe, C. J. *et al.* Mantle P-wave velocity structure beneath the Hawaiian hotspot. *Earth Planet. Sci. Lett.* **303**, 267–280 (2011).
24. Laske, G. *et al.* Asymmetric shallow mantle structure beneath the Hawaiian Swell—evidence from Rayleigh waves recorded by the PLUME network. *Geophys. J. Int.* **187**, 1725–1742 (2011).
25. Albers, M. & Christensen, U. The excess temperature of plumes rising from the core mantle boundary. *Geophys. Res. Lett.* **23**, 3567–3570 (1996).
26. Courtillot, V., Davaille, A., Besse, J. & Stutzmann, E. Three distinct types of hotspots in the Earth's mantle. *Earth Planet. Sci. Lett.* **205**, 295–308 (2003).
27. Steinberger, B. Plumes in a convecting mantle: models and observations for individual hotspots. *J. Geophys. Res.* **105**, 11127–11152 (2000).
28. Nataf, H. C. & VanDecar, J. Seismological detection of a mantle plume? *Nature* **364**, 115–120 (1993).
29. Su, W. J., Woodward, R. L. & Dziewonski, A. M. Degree 12 model of shear velocity heterogeneity in the mantle. *J. Geophys. Res.* **99**, 6945–6980 (1994).
30. Fukao, Y. & Obayashi, M. Subducted slabs stagnant above, penetrating through, and trapped below the 660 km discontinuity. *J. Geophys. Res.* **118**, 1–19 (2013).
31. Forte, A. M. & Mitrovica, J. K. Deep mantle high viscosity flow and thermochemical structure inferred from seismic and geodetic data. *Nature* **410**, 1049–1056 (2001).
32. Marquardt, H. & Miyagi, L. Slab stagnation in the shallow lower mantle due to an increase in mantle viscosity. *Nature Geosci.* **8**, 311–314 (2015).
33. Nolet, G., Karato, S. I. & Montelli, R. Plume fluxes from seismic tomography. *Earth Planet. Sci. Lett.* **248**, 685–699 (2006).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank the IRIS Data Management Center for providing the waveform data used in this study. This study was supported by an NSF Graduate Research Fellowship to S.W.F., NSF grant EAR-1417229, and ERC Advanced Grant WAVETOMO. Computations were performed at the National Energy Research Scientific Computing Center, supported by the US Department of Energy Office of Science (contract DE-AC02-05CH11231).

Author Contributions B.R. and S.W.F. collaborated in developing the concept of this paper. B.R. wrote the first draft, which was jointly finalized through successive iterations. S.W.F. is responsible for most of the technical aspects of this work, including the realization of most of the figures, with input from B.R. on figure design. B.R. is responsible for Fig. 4 and Extended Data Table 1.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to B.R. (barbara@seismo.berkeley.edu).

METHODS

Waveform inversion technique. As was the case for models SEMum³⁴ and SEMum2 (ref. 21), model SEMUCB-WM1⁴ was developed using a time-domain waveform inversion technique that combines highly accurate spectral-element forward wavefield modelling³⁵ with efficient sensitivity kernel computation using nonlinear asymptotic coupling theory (NACT)³⁶. This approach allows us to take advantage of both the accuracy of computing the misfit function with spectral-element modelling, and the efficiency of NACT-based kernels, which furthermore allow us to use a quickly converging Gauss–Newton scheme. Because the approximate Hessian matrix cannot easily fit in computer memory when considering whole-mantle scale inversions such as SEMUCB-WM1, we designed a high-performance distributed-memory abstraction³⁷ for its assembly from parallel NACT computations. As with other tomographic approaches, care must be taken in selecting the starting model owing to the strong nonlinearity of the inverse problem. In ref. 34, we explain how the present line of successive tomographic models (SEMum, SEMum2, and now SEMUCB-WM1) started from a one-dimensional earth model, and progressively incorporated more and more waveform data and shorter periods as iterations proceeded and three-dimensional structure became stronger.

Parameterization and starting model. Model SEMUCB-WM1 is constructed starting from model SEMum2²¹ above about 800-km depth and model SAW24B16³⁸ from below this depth (also used in the SEMum and SEMum2 inversions to account for lower-mantle structure). We invert for three-dimensional variations in Voigt-average isotropic shear-wave velocity (V_s) and the radially anisotropic parameter $\xi = (V_{sh}/V_{sv})^2$, with respect to a one-dimensional mean reference model that evolves throughout the inversion. In contrast, the one-dimensional attenuation model is fixed to a smoothed version of QL6 (ref. 39). We express perturbations to mantle V_s and ξ in 20 cubic b-splines³⁸ with variable spacing radially and in spherical splines⁴⁰ laterally, with lateral spacing of nodes of less than 2° for V_s and 8° for ξ . Source parameters for each event are kept fixed to those reported in the Global CMT catalogue (<http://www.globalcmt.org>). Tests indicate¹⁴ that allowing source perturbations should not measurably affect the resulting structure at the current resolution. From our SEMum2 + SAW24B16 starting model, three inversion iterations were performed, while incrementally incorporating more and shorter-period data (see ‘Data selection and inversion’). We note that the upper-mantle part of the model (down to 400 km) has not significantly changed in this process, further validating the results described in ref. 4.

Data selection and inversion. Our approach is described in detail in ref. 4 and includes discussions of how we calibrate our crustal model, introduce prior information in the data and model space, and assess model performance and uncertainties (including resolution and statistical resampling tests). Here we briefly recall details of the iterative inversion process, including the addition of new waveform data and tests of convergence.

While SEMum and SEMum2 included only fundamental and overtone mode surface waveforms at long periods ($T > 60$ s), here we include body-waveform data. To this end, we assembled a data set comprising full three-component teleseismic waveforms, filtered in multiple passbands, allowing us to incrementally incorporate higher-frequency body-waveform data: 1, surface-wave passband: cut-off at 400 s and 60 s (corners at 250 s and 80 s); 2, body-wave passband (filter I): cut-off at 300 s and 36 s (corners at 180 s and 45 s); and 3, body-wave passband (filter II): cut-off at 300 s and 32 s (corners at 180 s and 38 s). In addition to incorporating shorter periods, we also expanded our data set by incorporating additional events: starting from the 203 events used in developing upper-mantle models SEMum and SEMum2, we added 70 new events with moment magnitude $M_w = 5.8$ – 7.3 , chosen to be spatially distributed in a complementary manner to the original set.

The inversion comprised three phases. In phase I, we performed one iteration of inversion for whole-mantle structure using the 60-s surface-waveform and 36-s body waveform data sets (filter I) picked from the 203 events used in developing SEMum and SEMum2, and constraints from surface-wave group-velocity maps between 25 s and 150 s to enforce consistency with our crustal modelling scheme⁴. Because upper-mantle structure changed very little in this first iteration, aside from slightly larger amplitudes following the introduction of the body-waveform data, we chose to invert for structure only at depths greater than 300 km in the remaining iterations (after first recalibrating the crustal model one last time; see ref. 4 for details). In phase II, we introduced the 70 new events. This step involved picking the new-event data, as well as reprocessing the older-event data, using the spectral-element synthetics from the previous iteration (our waveform-data-picking approach selects data on the basis of their similarity to the spectral-element synthetics computed in the most recent iteration model). We then performed another inversion iteration, again using the 60-s and 36-s filter passbands, but now including the new-event data and inverting for structure below 300-km depth only. In phase III, we again reprocessed the data from the complete 273-event data

set, but now using a new shorter-period body-wave passband (filter II). We then inverted for structure below 300 km using the 60-s and 32-s data passbands.

To ensure that our inversion was converging, we determined, after each iteration, whether more waveform windows were selected in the subsequent data-reprocessing round than would have been selected using spectral-element synthetics from the previous-iteration model. By the final iteration, we found only small gains in the numbers of selected windows, indicating that the inversion had probably converged for the particular passbands considered. We also assessed convergence by testing fits to held-out waveforms from ten events not included in the inversion, and found that these validation data exhibited fits quite similar to that seen for the inversion data⁴.

Computational cost. For the four rounds of spectral-element simulations required to complete these three phases of inversion (including a last round to assess the final fit to the data), the present study required about three million CPU (central processing unit) hours, performed on Hopper, a Cray XE6 supercomputer at the National Energy Research Scientific Computing Center (NERSC). The NACT-based Hessian estimation and Gauss–Newton model update computations were performed on NERSC Edison, a Cray XC30.

Resolution and model uncertainties. It is well known that standard linear resolution analyses are strictly valid only for linear problems (or potentially near the optimum of a nonlinear problem⁴¹) and may also yield misleading results⁴². However, these remain useful techniques for assessing certain aspects of a tomographic inversions, providing insight into potential issues related to data coverage (for example, uneven sensitivity or smearing), the role of *a priori* information in constraining model smoothness, and limitations of the chosen model basis. We provide an extensive discussion of these aspects of SEMUCB-WM1 in the context of resolution analysis in ref. 4; here we focus on analyses specifically relevant to the types of conduit-like structures discussed in the main text.

Recovery of whole- and partial-mantle plumes. One obvious question that can be probed using resolution analysis is whether we can reasonably expect to resolve the plume-like conduits of the amplitude and scale observed in SEMUCB-WM1. In Extended Data Figs 4 and 5, we examine the recovery of synthetic whole- and partial-mantle plumes of diameter 1,000 km and 600 km, beneath Hawaii and Iceland. These synthetic test models have a peak amplitude of -2% , comparable to many of the conduits observed in our model, and a cosine-cap lateral amplitude profile (which means that the ‘core’ of each plume, exhibiting anomaly strength $>1\%$, is only half of its diameter—500 km and 300 km, respectively). We also examine recovery of plumes truncated at successively greater depths (1,000, 1,500 and 2,000 km) to assess vertical smearing. Artefacts above the truncation depth in the synthetic *i* models are due to aliasing phenomena associated with the radial b-spline basis used to parameterize our model (see above). Overall, we find that all whole- and partial-mantle input plumes are recovered quite well beneath both Hawaii, with denser data coverage, and Iceland, with comparatively sparser coverage (there is a slight difference in amplitude recovery beneath the two). We see no evidence of lateral smearing, or (in the case of the truncated plumes) radial smearing, nor do we detect notable gaps in recovery. Recovered amplitudes vary as a function of depth, with comparatively weaker, but still satisfactory, recovery in the less well sampled mid-mantle (about half of the input anomaly strength). Furthermore, as noted in the main text, the pattern of amplitude variation with depth seen in Extended Data Figs 4 and 5 does not match that of our imaged plumes, which often show local amplitude maxima in the mid-mantle (not minima, as suggested by these tests). Thus, while we cannot rule out that spatial variation in sensitivity contributes in some way to the imaged amplitude distribution, these results give us confidence that plumes of similar dimension and amplitude to those seen in SEMUCB-WM1 should be recoverable.

In Extended Data Fig. 6, we consider a narrower plume (400 km in diameter) spanning from the CMB to 1,000-km depth, with a peak amplitude of -2% . We observe that the output structure is at least 800 km in width, but is also significantly weaker than the input, exhibiting a maximum amplitude of -0.6% near its base, while only reaching -0.3% or -0.4% elsewhere in its core, indicating that an actual plume of width less than 400 km would have to be much stronger to be properly detected in our model (see further discussion in ‘Estimated excess temperatures and actual width of plumes’).

Recovery of ‘hanging’ plume structures. To supplement these analyses, we also explored tests using synthetic ‘hanging’ plumes—columnar anomalies extending down from the surface into the upper mantle and transition zone. These plumes have a diameter of 600 km and again a cosine-cap cross-section, as well as -2% maximum amplitude, but are now truncated at depths of 410 km or 1,000 km (Extended Data Figs 7 and 8). This experiment is designed to further assess the effect of depth smearing, and again examines two geographic locations: Hawaii, with denser, and Iceland, with sparser, data coverage. In general, the retrieved output structures are remarkably symmetrical and exhibit the correct depth extent, with the exception of the plume truncated at 1,000-km depth beneath Hawaii,

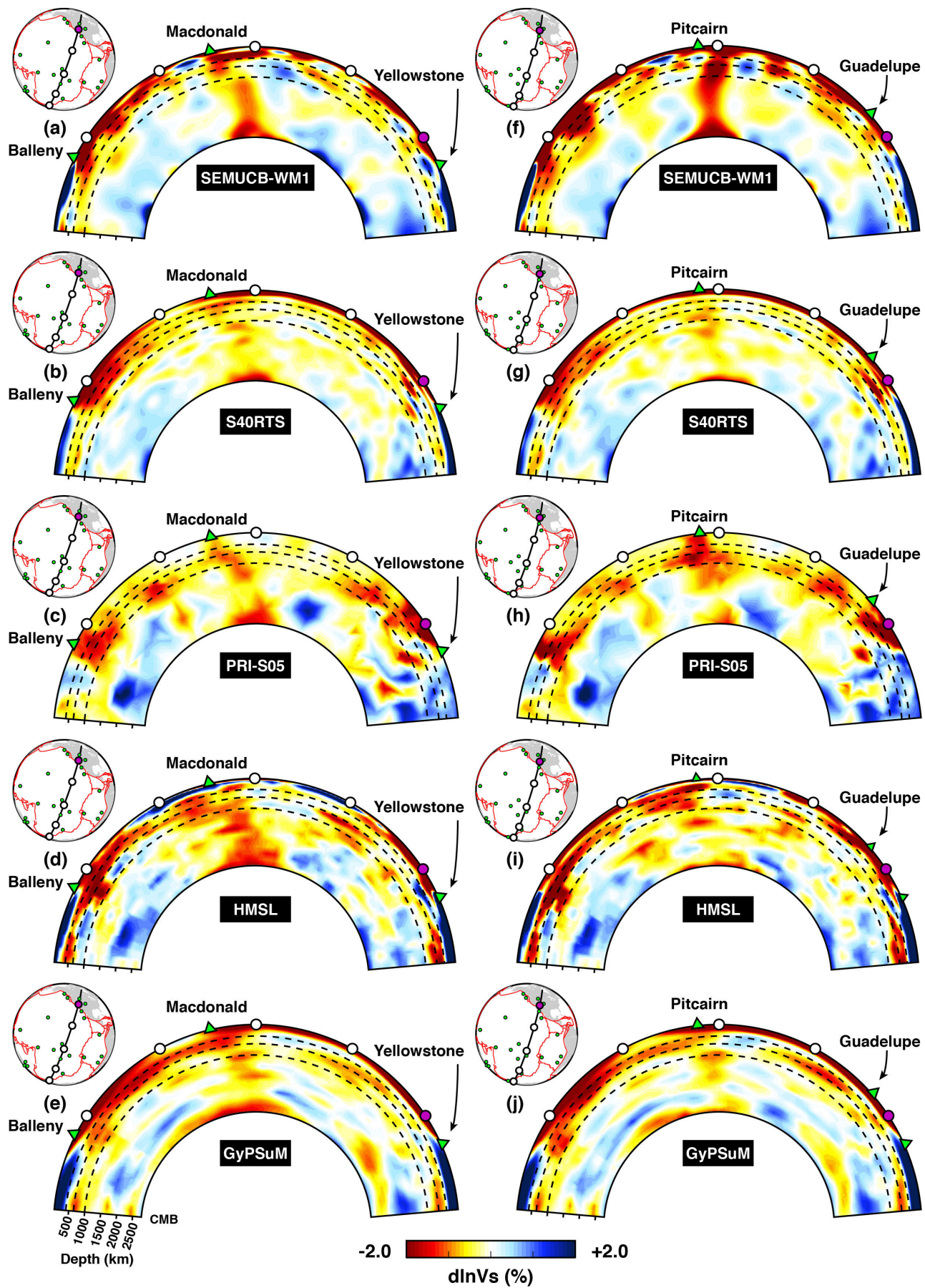
which shows a slight eastward-trending band extending to the CMB. We note that this artefact is very weak, everywhere less than 0.1% amplitude (that is, at least 20× weaker than the −2% input amplitude). Furthermore, this artefact in no way resembles the plume we image beneath Hawaii in SEMUCB-WM1; it has a very different trend and amplitude profile. Turning now to the results beneath Iceland, we again see that retrieved amplitudes are comparatively weaker, presumably because of the sparser data coverage, but at the same time we find no anomalous band. This observation suggests that while Hawaii is expected to have denser data coverage, the data coverage beneath Hawaii might also be more anisotropically distributed. Indeed, this would be consistent with the eastward or east–southeast trending streaks seen at mid-mantle depths in some travel-time-based local tomographic studies in this region. Overall, these results give us further confidence that the depth extensions of the plumes we image in SEMUCB-WM1 cannot be attributed to smearing.

Estimated excess temperatures and actual width of plumes. Our resolution tests indicate that we can easily resolve a synthetic columnar velocity anomaly over a width of 1,000 km and maximum amplitude of 2% (Extended Data Fig. 4). This is indeed comparable to the maximum amplitudes of the plume conduits imaged in the mid- and lower mantle in our model. If the actual width of a plume is instead significantly smaller than 1,000 km, then the average velocity anomaly should be correspondingly larger to attain the same imaged amplitudes. For example, if the actual plume diameter is 600 km (Extended Data Fig. 5), then the actual velocity anomaly would have to be on the order of 4%–5% (as the recovered amplitudes in the 600-km-width case are on the order of ≤ 0.5 times those obtained for the 1,000-km-width case). Similarly, an even narrower plume, with a diameter of 400 km or less (Extended Data Fig. 6), should require a velocity contrast within the plume of more than 10% to be detected in our inversion.

Assuming the velocity anomaly is due to temperature alone, a 2% increase in V_s translates into about 200 K excess temperature (ΔT) in the upper mantle^{43,50}. However, partial derivatives of shear velocity with respect to temperature greatly decrease with pressure (that is, with depth in the mantle). While they are not precisely known at lower mantle conditions, a factor-of-two reduction at mid-to-lower mantle depths is a reasonable assumption^{44,51}, translating a 2% increase in V_s into $\Delta T \approx 400$ K, and 10% into a most unrealistic ΔT in excess of 2,000 K. Thus, assuming that the relative amplitude recovery in our resolution analyses is representative of reality, and that plumes are purely thermal, it is far more plausible that we are correctly resolving broader weaker plumes in the lower mantle than poorly resolving very strong narrower ones.

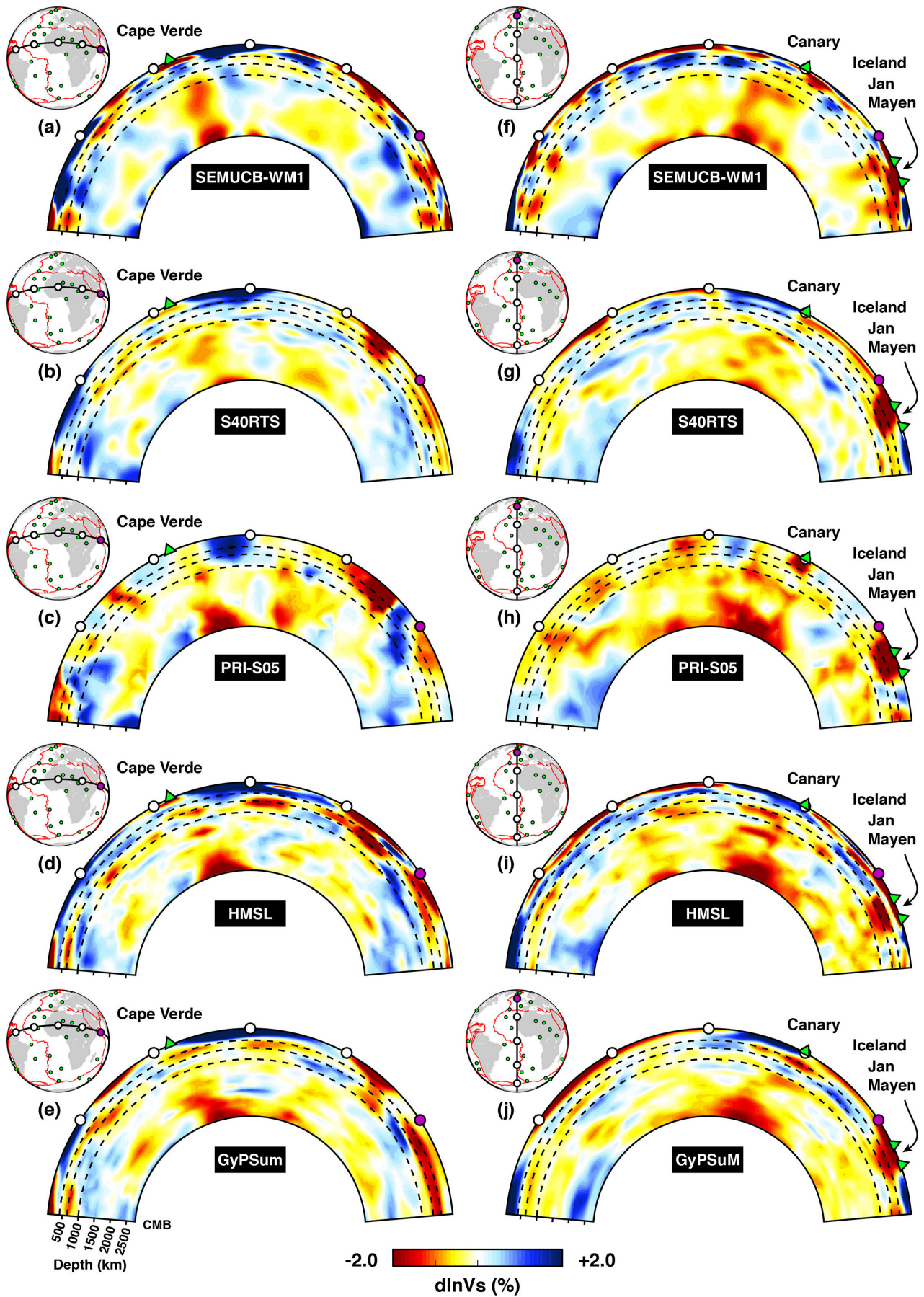
34. Lekić, V. & Romanowicz, B. Inferring upper-mantle structure by full waveform tomography with the spectral element method. *Geophys. J. Int.* **185**, 799–831 (2011).

35. Komatitsch, D. & Vilotte, J. P. The spectral element method: an efficient tool to simulate the seismic response of 2D and 3D geological structures. *Bull. Seismol. Soc. Am.* **88**, 368–393 (1998).
36. Li, X.-D. & Romanowicz, B. Comparison of global waveform inversions with and without considering cross-branch modal coupling. *Geophys. J. Int.* **121**, 695–709 (1995).
37. French, S. W., Zheng, Y., Romanowicz, B. & Yelick, K. Parallel Hessian assembly for seismic waveform inversion using global updates. In *Proc. 29th IEEE Int. 'Parallel and Distributed Processing' Symp.* (IEEE, <http://dx.doi.org/10.1109/IPDPS.2015.58> (2015)).
38. Mégnin, C. & Romanowicz, B. The three-dimensional shear velocity structure of the mantle from the inversion of body, surface and higher-mode waveforms. *Geophys. J. Int.* **143**, 709–728 (2000).
39. Durek, J. J. & Ekström, G. A radial model of anelasticity consistent with long-period surface-wave attenuation. *Bull. Seismol. Soc. Am.* **86**, 144–158 (1996).
40. Wang, Z. & Dahlen, F. A. Spherical-spline parametrization of three-dimensional earth models. *Geophys. Res. Lett.* **22**, 3099–3102 (1995).
41. Tarantola, A. *Inverse Problem Theory and Models for Model Parameter Estimation* (Society for Industrial and Applied Mathematics (SIAM), 2005).
42. Lévêque, J., Rivera, L. & Wittlinger, G. On the use of the checker-board test to assess the resolution of tomographic inversions. *Geophys. J. Int.* **115**, 313–318 (1993).
43. Macpherson, C. G., Hilton, D. R., Sinton, J. M., Poreda, R. J. & Craig, H. High $^3\text{He}/^4\text{He}$ ratios in the Manus backarc basin: implications for mantle mixing and the origin of plumes in the western Pacific Ocean. *Geology* **26**, 1007–1010 (1998).
44. Ritsema, J., van Heijst, H. J. & Woodhouse, J. H. Complex shear wave velocity structure imaged beneath Africa and Iceland. *Science* **286**, 1925–1928 (1999).
45. Kreemer, C. Absolute plate motions constrained by shear wave splitting orientations with implications for hot spot motions and mantle flow. *J. Geophys. Res.* **114**, B10405 (2009).
46. Ritsema, J., Deuss, A., Van Heijst, H. & Woodhouse, J. S40RTS: a degree-40 shear-velocity model for the mantle from new Rayleigh wave dispersion, teleseismic traveltime and normal-mode splitting function measurements. *Geophys. J. Int.* **184**, 1223–1236 (2011).
47. Montelli, R., Nolet, G., Dahlen, F. A. & Masters, G. A catalogue of deep mantle plumes: new results from finite-frequency tomography. *Geochim. Geophys. Geosyst.* **7**, Q11007 (2006).
48. Houser, C., Masters, G., Shearer, P. & Laske, G. Shear and compressional velocity models of the mantle from cluster analysis of long-period waveforms. *Geophys. J. Int.* **174**, 195–212 (2008).
49. Simmons, N. A., Forte, A. M., Boschi, L. & Grand, S. P. GyPSuM: a joint tomographic model of mantle density and seismic wave speeds. *J. Geophys. Res.* **115**, B12310 (2010).
50. Stixrude, L. & Lithgow-Bertelloni, C. Mineralogy and elasticity of the oceanic upper mantle: Origin of the low-velocity zone. *J. Geophys. Res.* **110**, B03204 (2005).
51. Karato, S.-I. *Deformation of Earth Materials: Introduction to the Rheology of the Solid Earth* (Cambridge Univ. Press., 2008).



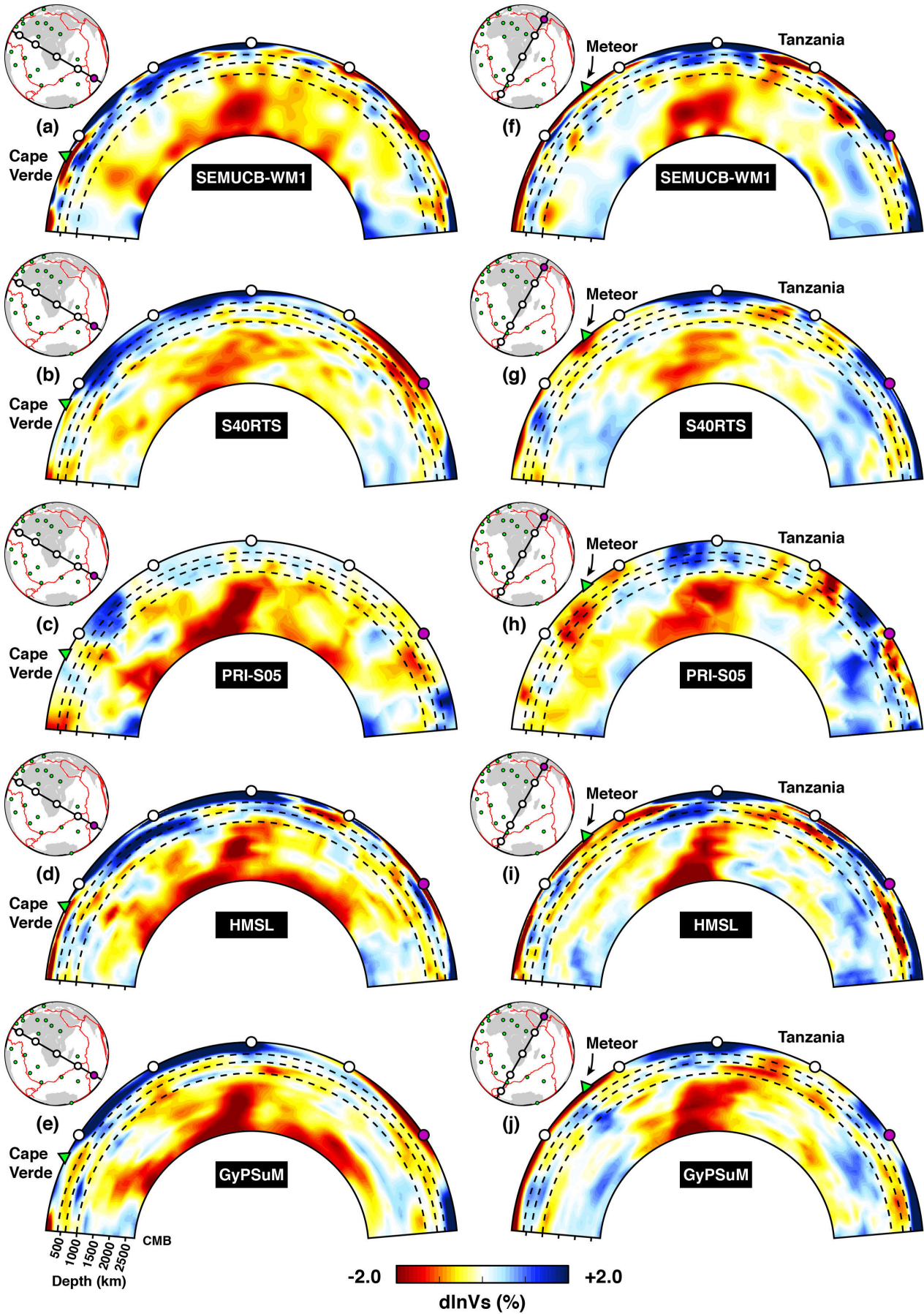
Extended Data Figure 1 | Intermodel comparisons. These figures correspond to the cross-sections in Fig. 1c and d, oriented normal to the direction of Pacific absolute plate motion⁴⁵. As in Fig. 1, sections are indicated in the inset maps, while white and purple circles indicate position along section and orientation. Shown are relative shear-wave velocity (V_s) anomalies in models SEMUCB-WM1 (this study), S40RTS (ref. 46), PRI-S05 (ref. 47), HMSL-S06 (ref. 48) and GyPSuM (ref. 49), each plotted with respect to its own one-dimensional reference (where the latter notion is well defined: see for example ref. 48; where defined, the one-dimensional reference is often the global

average). Panels a–e correspond to the MacDonald-hotspot-centred view of Fig. 1d; panels f–j correspond to the Pitcairn-centred view of Fig. 1c. This comparison shows that the five models are broadly compatible with each other at long wavelengths. However, in the lower mantle, the MacDonald and Pitcairn plumes are much more clearly defined as vertical conduits in SEMUCB-WM1, and stand out as the strongest and most continuous low-velocity features in the lower mantle in these cross-sections (which span almost half of Earth's circumference).



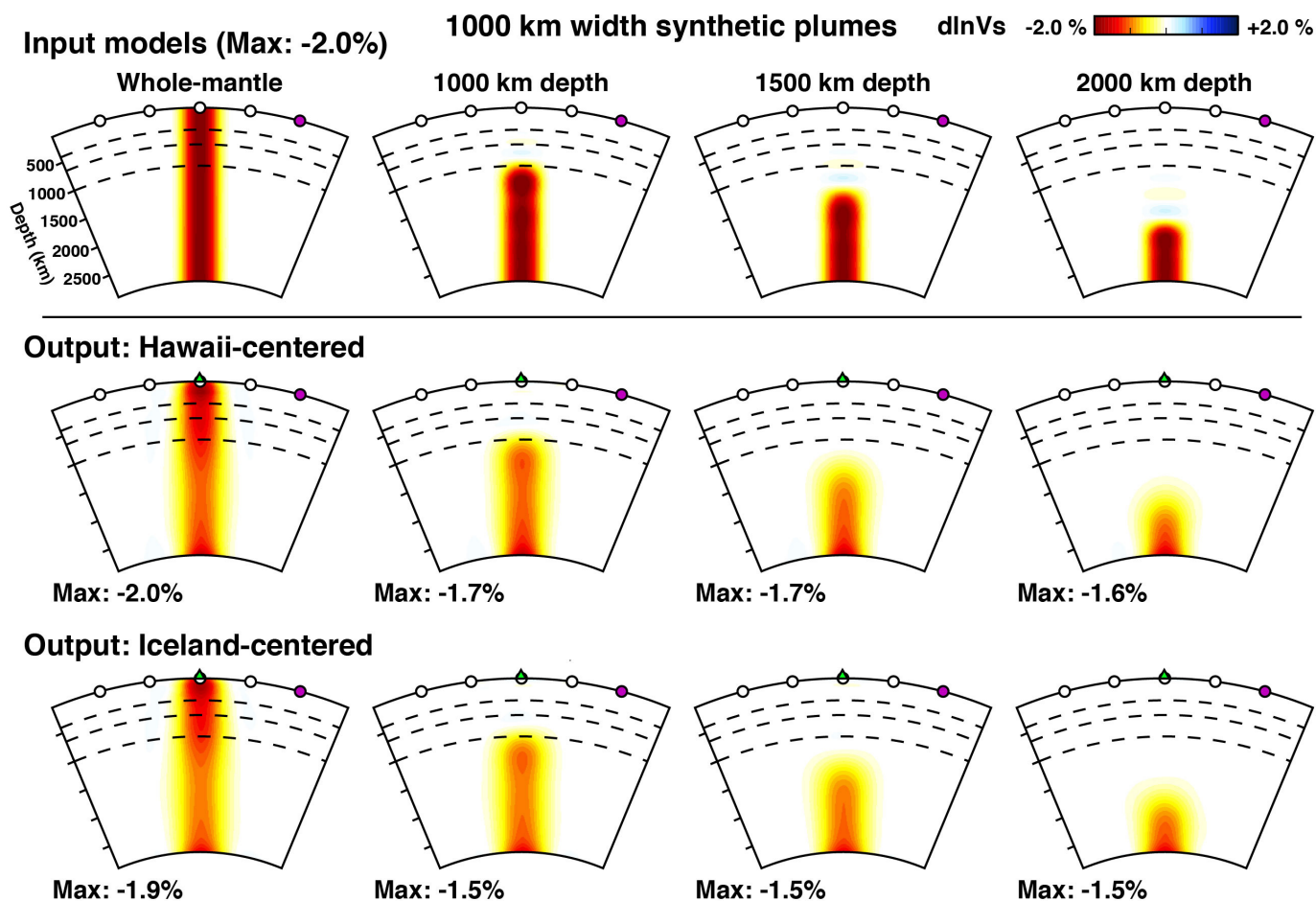
Extended Data Figure 2 | Intermodel comparisons. These comparisons correspond to Fig. 1e, f, presented in a similar manner to those in Extended Data Fig. 1. We again find that models SEMUCB-WM1, S40RTS, PRI-S05, HMSL-S06 and GyPSuM are broadly compatible with each other at long wavelengths. However, in the lower mantle, the plumes beneath both Cape Verde and Canary are more clearly defined as well isolated vertical conduits in

SEMUCB-WM1. Furthermore, while we do observe some degree of correspondence between the two plumes imaged in SEMUCB-WM1 and some anomalies also present in PRI-S05 or HMSL (for example, the plume root at the CMB beneath Cape Verde, or the lateral translation of the plume around 1,000 km beneath Canary), the unambiguously columnar nature of the anomalies imaged in SEMUCB-WM1 stands in stark contrast.



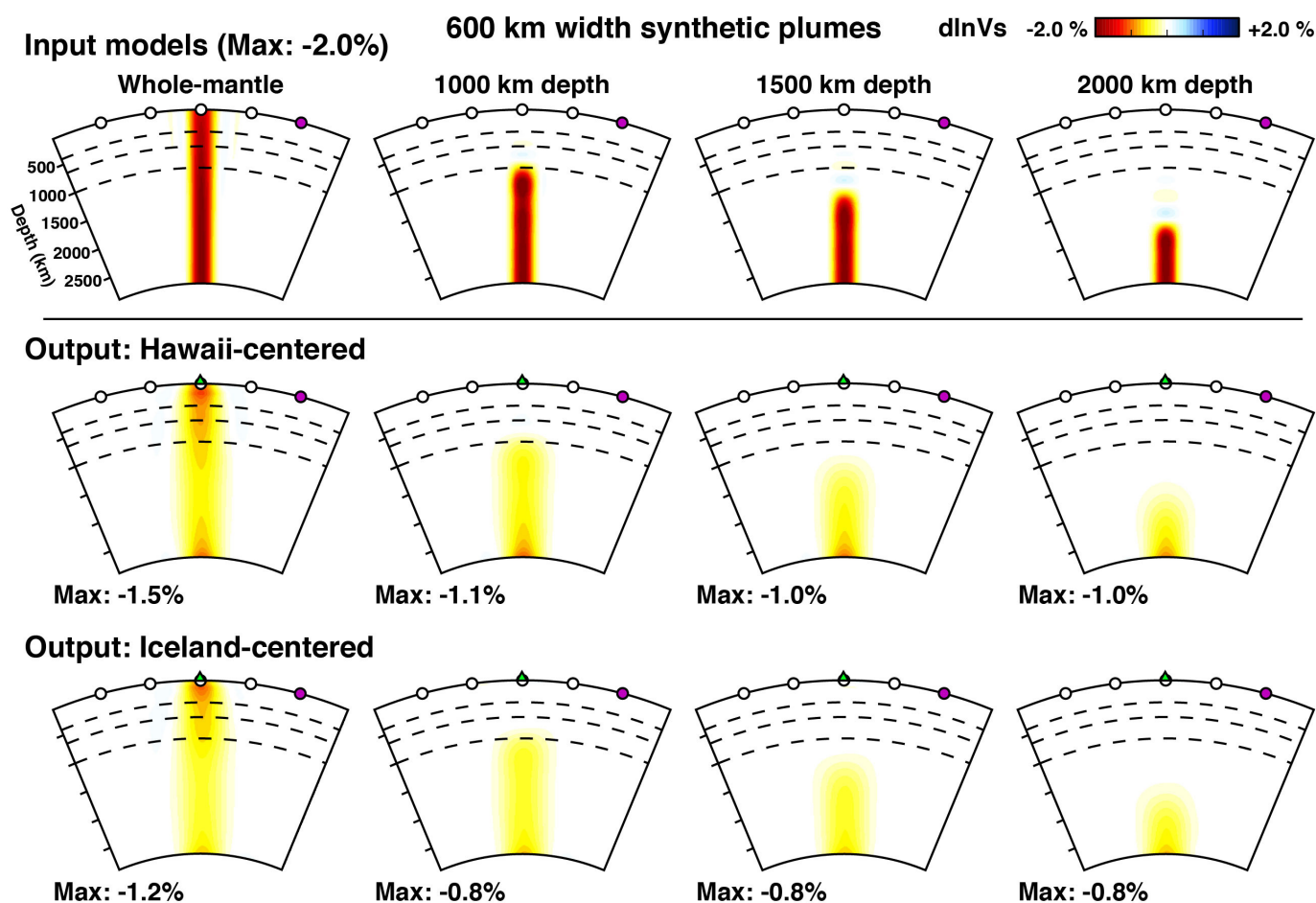
Extended Data Figure 3 | Inter-model comparisons. These cross-sections are similar to those in Extended Data Fig. 1, but now feature two approximately orthogonal sections through the African LLSVP: **a–e**, traversing from northwest to southeast; **f–j**, traversing from southwest to northeast. The

African LLSVP is more massive, and therefore better resolved in S40RTS, PRI-S05, HMSL-S06 and GyPSuM than are other plumes. As in Extended Data Fig. 2, we again note some degree of similarity between SEMUCB-WM1 (**a**), PRI-S05 (**c**) and HMSL-S06 (**d**) below the Cape Verde plume.



Extended Data Figure 4 | Linear resolution analysis, examining recovery of synthetic whole- and partial-mantle plumes of width 1,000 km beneath Hawaii and Iceland. Synthetic plume input models, shown in the upper row of panels, have a peak amplitude of -2% and a cosine-cap lateral amplitude profile (thus, the effective width above 1% anomaly strength is only 500 km). In addition to looking at a whole-mantle plume, we also examine recovery of plumes truncated at successively greater depths (1,000 km, 1,500 km and 2,000 km) to assess vertical smearing. Artefacts seen above the truncation depth in the synthetic input models are due to slight aliasing phenomena associated with the radial b-spline basis functions used to parameterize our model. We find that all four input plumes are recovered quite well beneath both Hawaii

(centre row), with relatively denser data coverage, and Iceland (bottom row), with comparatively sparser coverage—although there is a slight difference in amplitude recovery beneath the two (maximum amplitude recovered is shown for each panel). Importantly, we see no evidence of lateral (or, in the case of the truncated plumes, radial) smearing, nor do we detect significant gaps in recovery. However, recovered amplitude does vary as a function of depth, with comparatively weaker, although still satisfactory, recovery in the less well sampled mid-mantle (of the order of half of the input anomaly strength). For a more thorough discussion of the caveats implied by linear resolution analysis in the context of our inversion, as well as additional resolution tests, see ref. 4.

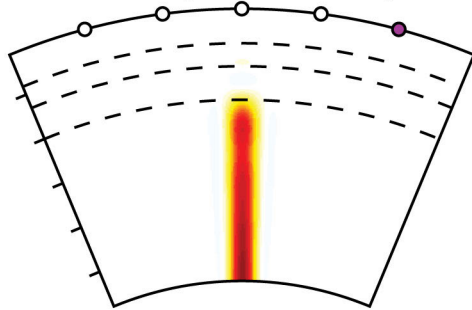


Extended Data Figure 5 | Linear resolution analysis. Similar to that in Extended Data Fig. 4. This analysis again features whole- and partial-mantle plumes of peak strength -2% , but now of width 600 km (meaning the effective width above 1% anomaly strength is only 300 km). We again find that the synthetic input plumes are recovered quite well, with no evidence of lateral or radial smearing, as well as no gaps in recovery. At the same time, we find that recovered amplitude is poorer than for the larger, 1,000-km-width plumes,

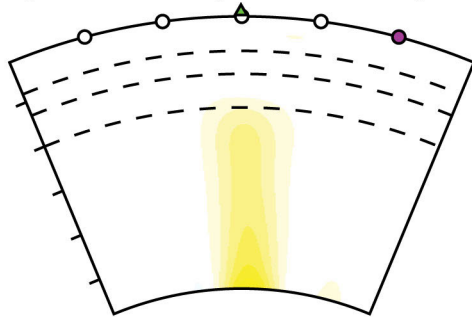
in some cases recovering amplitudes of the order of one-quarter of the input, and we note that there is again a slight disparity in amplitude recovery between Hawaii and Iceland. Furthermore, we note that tests using synthetic plumes at or below widths of 600 km push the limits of the spherical-spline lateral basis functions used in our model—particularly in the upper mantle, where the inter-spline absolute distance is larger (although the angular distance remains constant).


400 km width synthetic plume

Input model (Max: -2.0%)

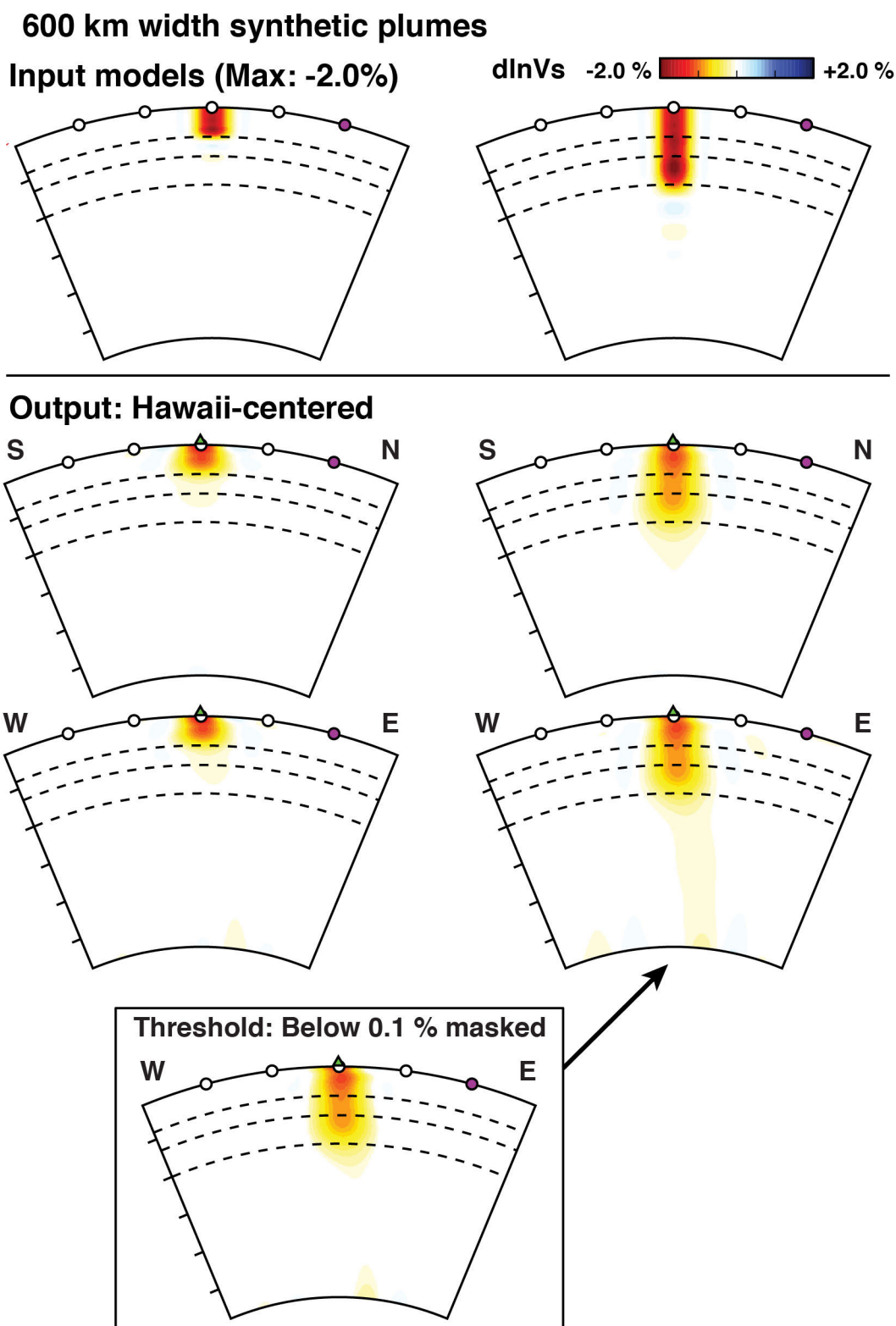


Output model (Max: -0.6%)



dlnVs -2.0 %  +2.0 %

Extended Data Figure 6 | Further linear resolution analysis. This analysis is of a 400-km-width plume-like conduit extending from the CMB to 1,000-km depth, with a similar lateral profile (a cosine cap) and maximum amplitude (-2%) as the test structures in Extended Data Figs 4 and 5. The inherent limits of our spherical spline basis prohibit us from representing this narrow conduit with sufficient fidelity for the purposes of this test above 1,000 km. Upper panel, conduit-like input structure; lower panel, output structure resulting from resolution test. We observe that the output structure is at least 800 km in width, but is also significantly weaker than the input, exhibiting a maximum amplitude of -0.6% near its base, while only reaching -0.3 or -0.4% elsewhere in its core. As such, we can infer that the input-structure amplitudes would need to be increased by at least $10\times$ in order to maintain amplitudes near -2.0% throughout the majority of the lower mantle. This latter observation has implications for effective excess temperature (see Methods).




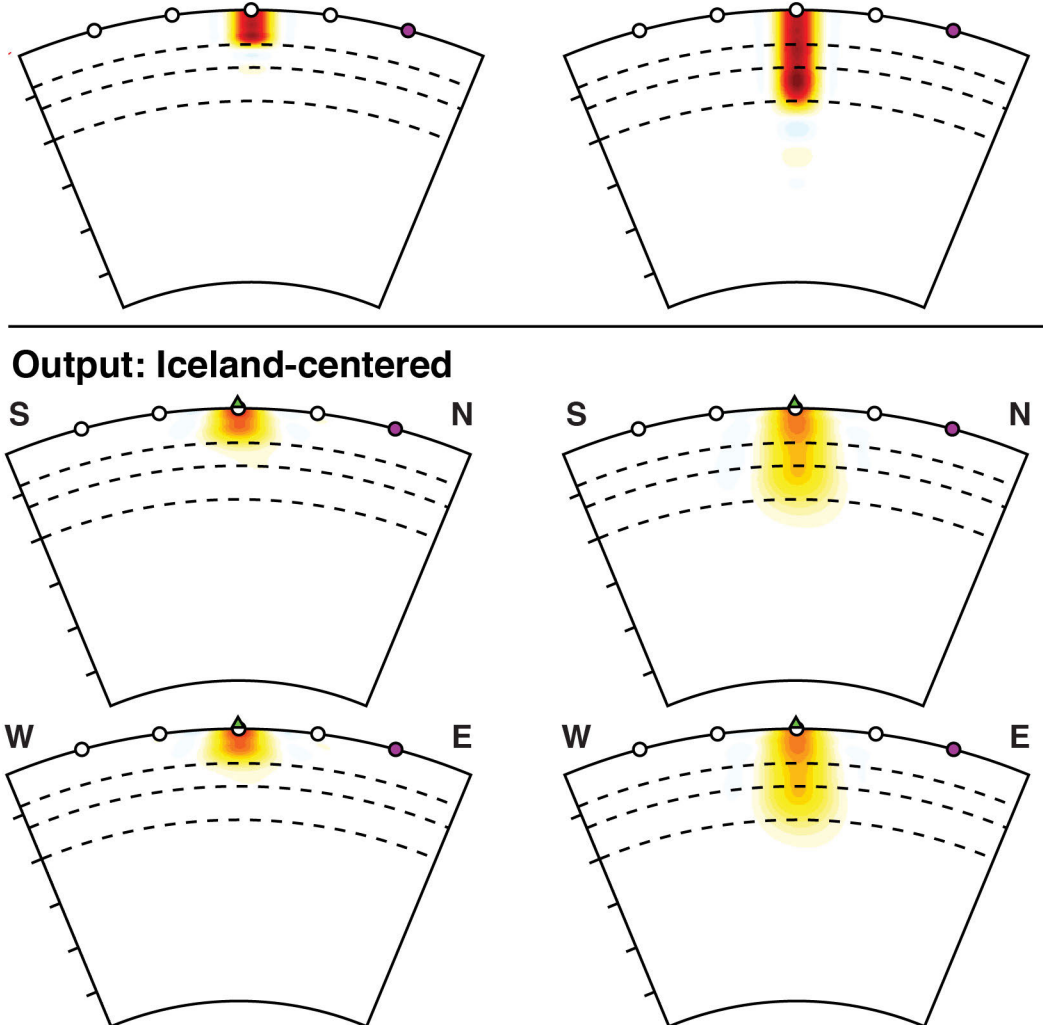
Extended Data Figure 7 | Linear resolution analysis for synthetic ‘hanging’ plume input structures in the upper mantle and transition zone. Like those in Extended Data Figs 4 and 5, these plumes have an overall width of 600 km and a cosine-cap lateral cross-section, as well as -2% maximum amplitude, but are now cut at 410-km (left panels) or 1,000-km (right panels) depth. This experiment is designed to assess the effect of depth-smearing in SEMUCB-WM1. Upper panels, hanging-plume input models. Lower panels, output models when inputs are placed beneath Hawaii. We note that in general

the structures retrieved are quite symmetrical and exhibit the appropriate depth extent, with the exception of the plume truncated at 1,000 km, which shows a weak eastward-trending band extending to the CMB. We note that this artefact is very weak, generally less than 0.1% amplitude, as illustrated in the bottom panel, where structure below 0.1% is masked (that is, the band is at least $20\times$ weaker than the -2% input structure). Furthermore, we note that this feature is not at all like the plume we image beneath Hawaii, as it possesses a very different trend and amplitude profile.

600 km width synthetic plumes

Input models (Max: -2.0%)

dlnVs -2.0 %  +2.0 %



Extended Data Figure 8 | Linear resolution analysis for synthetic ‘hanging’ plume input structures in the upper mantle and transition zone. This figure is similar to Extended Data Fig. 7, but now examines recovery beneath Iceland. Upper panels, hanging-plume input models. Lower panels, output

models when inputs are placed beneath Iceland. The retrieved structures are again quite symmetrical and exhibit the appropriate depth extent, although amplitude recovery is slightly less impressive than that observed beneath Hawaii (consistent with the results of Extended Data Figs 4 and 5).

Extended Data Table 1 | Plumes detected in the lower mantle in model SEMUCB-WM1⁴, and corresponding hotspots

Index of plume (Fig. 4)	Hotspot name	Ranking by Courtillot ²⁶	Buoyancy flux ²⁶	³ He/ ⁴ He ²⁶
Category 1: "Primary Plume"				
1	Afar	4	1	high
2	Canary	2	1	low
3	Cape Verde	2	1.6	high
4	Comores	0+	?	?
5	Hawaii	4+	8.7	high
6	Iceland	4+	1.4	high
7	Macdonald	2+	3.3	high?
8	Marqueses	2+	3.3	low
9	Pitcairn	2+	3.3	high?
10	Samoa	4	1.6	high
11	Tahiti/Society	2+	3.3	high?
Category 2: "Clearly resolved"				
12	Cameroon	0+	?	?
13	Caroline	3	2	high
14	Easter	4+	3	high
15	Galapagos	2+	1	high
16	Louisville	3+	0.9	?
17	Reunion	4	1.9	high
18	St Helena	1	0.5	low
19	Tristan	3	1.7	low
20	Kerguelen	2+	0.5	high
Category 3: "Somewhat resolved"				
21	Ascension	0+	1	?
22	Azores	1+	1.1	high?
23	Bouvet	1+	0.4	high
24	Crozet/Pr.Edw/	0+	0.5	?
25	Hoggar	1	0.9	?
26	Juan Fernandez	2+	1.6	high
27	San Felix	1+	1.6	?
Not associated with any known hotspot				
	Indonesia			

The numbering (column 1) and categories correspond to those in Fig. 4. Plumes are categorized as primary if the corresponding low-velocity conduit in the lower mantle has $\delta V_s/V_s$ less than -1.5% for most of the depth interval 1,000–2,800 km. These 11 plumes also correspond to regions of the lower mantle where the average velocity reduction over the depth range 1,000–1,800 km is significant at the 2σ level (see, for example, Supplementary Figs 3 and 4). Clearly resolved plumes correspond to vertically continuous conduits with $\delta V_s/V_s < -0.5\%$ in the depth range 1,000–2,800 km. Somewhat resolved plumes have vertically trending conduits with $\delta V_s/V_s < -0.5\%$ for most of the depth range 1,000–2,800 km, albeit not as clearly continuous. The only clearly resolved plume in the lower mantle that is not near a hotspot is in Indonesia, possibly because it is rising beneath a broad slab. However, it occurs close to a location where high $^3\text{He}/^4\text{He}$ ratios have been observed⁴³. For comparison, we list the corresponding hotspot ranking (column 3)²⁶, as well as the buoyancy flux (column 4) and $^3\text{He}/^4\text{He}$ ratios (column 5)²⁶. Question marks indicate no value given in ref. 26. Note that in this previous ranking of hotspots, these estimates of buoyancy flux and $^3\text{He}/^4\text{He}$ ratios were used together with the velocity anomaly values in the transition zone (500-km depth) from an older tomographic shear-velocity model⁴⁴. In contrast, our ranking is based entirely on the continuity of broad vertically oriented low-velocity structures across the major part of the lower mantle. Hotspots that do not have any clear expression in the lower mantle in model SEMUCB-WM1 are not listed, namely Yellowstone, Juan de Fuca/Cobb and Bowie (see also Fig. 4).

Global exchange and accumulation of non-native plants

Mark van Kleunen¹, Wayne Dawson¹, Franz Essl², Jan Pergl³, Marten Winter⁴, Ewald Weber⁵, Holger Kreft⁶, Patrick Weigelt⁶, John Kartesz⁷, Misako Nishino⁷, Liubov A. Antonova⁸, Julie F. Barcelona⁹, Francisco J. Cabezas¹⁰, Dairon Cárdenas¹¹, Juliana Cárdenas-Toro^{12,13}, Nicolás Castaño¹¹, Eduardo Chacón^{2,14}, Cyrille Chatelain¹⁵, Aleksandr L. Ebel¹⁶, Estrela Figueiredo^{17,18}, Nicol Fuentes¹⁹, Quentin J. Groom²⁰, Lesley Henderson²¹, Inderjit²², Andrey Kupriyanov²³, Silvana Masciadri^{24,25}, Jan Meerman²⁶, Olga Morozova²⁷, Dietmar Moser²⁸, Daniel L. Nickrent²⁸, Annette Patzelt²⁹, Pieter B. Pelsers⁹, María P. Baptiste¹², Manop Poopath³⁰, Maria Schulze³¹, Hanno Seebens³², Wen-sheng Shu³³, Jacob Thomas³⁴, Mauricio Velazco¹⁰, Jan J. Wieringa^{35,36} & Petr Pyšek^{3,37,38}

All around the globe, humans have greatly altered the abiotic and biotic environment with ever-increasing speed. One defining feature of the Anthropocene epoch^{1,2} is the erosion of biogeographical barriers by human-mediated dispersal of species into new regions, where they can naturalize and cause ecological, economic and social damage³. So far, no comprehensive analysis of the global accumulation and exchange of alien plant species between continents has been performed, primarily because of a lack of data. Here we bridge this knowledge gap by using a unique global database on the occurrences of naturalized alien plant species in 481 mainland and 362 island regions. In total, 13,168 plant species, corresponding to 3.9% of the extant global vascular flora, or approximately the size of the native European flora, have become naturalized somewhere on the globe as a result of human activity. North America has accumulated the largest number of naturalized species, whereas the Pacific Islands show the fastest increase in species numbers with respect to their land area. Continents in the Northern Hemisphere have been the major donors of naturalized alien species to all other continents. Our results quantify for the first time the extent of plant naturalizations worldwide, and illustrate the urgent need for globally integrated efforts to control, manage and understand the spread of alien species.

The magnitude of impacts caused by alien species on native biota and human societies is increasing rapidly³. However, our knowledge of the global spread and distribution of naturalized species (that is, alien species that form self-sustaining populations in new regions^{4,5}) is still very limited. Nevertheless, there are many presumptions about the distributions and patterns of spread of alien species. For example, it has frequently been suggested that Old World species have spread more widely outside their native ranges than New World species, owing to human colonization history or intrinsic evolutionary

superiority^{6,7}. It has also been suggested that islands have more alien species than mainland areas, among others because of unfilled niche space on islands^{7,8} or, as shown for birds, a higher introduction effort⁹. Although these hypotheses have been tested for some parts of the world^{9,10}, global tests are still lacking.

Scientific and societal concerns about alien species have led to improved documentation of their distributions, and inventories have become available for many regions¹¹. Many of these inventories are still incomplete, especially for megadiverse taxonomic groups that are difficult to survey, such as invertebrates and microorganisms, and for less well-surveyed regions. However, vascular plants are well documented because of long histories of exploration. Recently, there have been several major efforts to combine inventories of alien species for large geographical regions (for example, Delivering Alien Invasive Species Inventories for Europe (DAISIE; <http://www.europe-alien.org/>)) and for those considered to be the most problematic invaders globally¹². However, a global database of the distribution of all naturalized alien plant species had not yet been built. Such data are essential for understanding global naturalization patterns and their underlying processes, reporting biodiversity status in terms of essential biodiversity variables¹³, and informing environmental managers across political borders via early warning systems.

Here, we present an analysis of naturalized vascular plant species in 843 non-overlapping regions (countries, federal states, islands) covering ~83% of the Earth's land surface (Fig. 1). We used a novel database, Global Naturalized Alien Flora (GloNAF), combined with data on the origins of the naturalized species and estimates of the numbers of native species per continent, to assess (1) which continents have accumulated the largest naturalized floras, and (2) which have been the major donors of naturalized alien plant species to other parts of the world.

¹Ecology, Department of Biology, University of Konstanz, Universitätsstrasse 10, D-78464 Konstanz, Germany. ²Division of Conservation, Vegetation and Landscape Ecology, University of Vienna, 1030 Wien, Austria. ³Institute of Botany, Department of Invasion Ecology, The Czech Academy of Sciences, CZ-252 43 Průhonice, Czech Republic. ⁴German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Deutscher Platz 5e, D-04103 Leipzig, Germany. ⁵Institute of Biochemistry and Biology, University of Potsdam, D-14469 Potsdam, Germany. ⁶Biodiversity, Macroecology & Conservation Biogeography, University of Göttingen, Büsgenweg 1, D-37077 Göttingen, Germany. ⁷Biota of North America Program (BONAP), Chapel Hill, North Carolina 27516, USA. ⁸Institute for Aquatic and Ecological Problems, Far East Branch, Russian Academy of Sciences, 680000 Khabarovsk, Russia. ⁹School of Biological Sciences, University of Canterbury, Private Bag 4800, Christchurch 8140, New Zealand. ¹⁰Departamento de Biodiversidad y Conservación, Real Jardín Botánico, CSIC, Plaza de Murillo 2, 28014 Madrid, Spain. ¹¹Instituto Amazónico de Investigaciones Científicas Sinchi, Herbario Amazónico Colombiano, 110311 Bogotá, Colombia. ¹²Instituto de Investigación de Recursos Biológicos Alexander von Humboldt, 111311 Bogotá, Colombia. ¹³Arts Faculty, Monash University, 3145 Melbourne, Australia. ¹⁴Escuela de Biología, Universidad de Costa Rica, 11501 San José, Costa Rica. ¹⁵Conservatoire et jardin botaniques de la Ville de Genève, 1292 Genève, Switzerland. ¹⁶Laboratory of Plant Taxonomy and Phylogeny, Tomsk State University, Lenin Prospect 36, 634050, Tomsk, Russia. ¹⁷Department of Botany, PO Box 77000, Nelson Mandela Metropolitan University, Port Elizabeth, 6031 South Africa. ¹⁸Centre for Functional Ecology, Departamento de Ciências da Vida, Universidade de Coimbra, 3001-455 Coimbra, Portugal. ¹⁹Facultad de Ciencias Forestales, Instituto de Ecología y Biodiversidad, Universidad de Concepción, Victoria 631, 403000, Concepción, Chile. ²⁰Botanic Garden Meise, Domein van Bouchout, B-1860, Meise, Belgium. ²¹ARC-Plant Protection Research Institute, Pretoria 0001, South Africa. ²²Department of Environmental Studies and Centre for Environmental Management Degraded of Ecosystems, University of Delhi, Delhi 110007, India. ²³Institute of Human Ecology SB RAS, Pr. Leningradskiy 10, 650065 Kemerovo, Russia. ²⁴Programa de Pós-graduação em Ecologia, UFRN, Campus Lagoa Nova, 59078-900 Natal, Brazil. ²⁵Oceanología y Ecología Marina, Facultad de Ciencias, Universidad de la República, Iguá, 4225, CP 11400, Montevideo, Uruguay. ²⁶Belize Tropical Forest Studies, PO Box 208, Belmopan, Belize. ²⁷Institute of Geography RAS, Staromonetny, 29, 119017 Moscow, Russia. ²⁸Department of Plant Biology, Southern Illinois University, Carbondale, Illinois 62901-6509 USA. ²⁹Oman Botanic Garden, Diwan of Royal Court, 122 Muscat, Oman. ³⁰The Forest Herbarium (BKF), Department of National Parks, Wildlife and Plant Conservation, Chatuchak, Bangkok 10900, Thailand. ³¹Department of Biology, Martin-Luther University Halle-Wittenberg, D-06108 Halle, Germany. ³²Institute for Chemistry and Biology of the Marine Environment, University of Oldenburg, Carl-von-Ossietzky Straße 9–11, D-26111 Oldenburg, Germany. ³³State Key Laboratory of Biocontrol and Guangdong Key Laboratory of Plant Resources, College of Ecology and Evolution, Sun Yat-sen University, Guangzhou 510275, China. ³⁴Department of Botany & Microbiology, College of Science, King Saud University, PO Box 2455, Riyadh 11451, Saudi Arabia. ³⁵Naturalis Biodiversity Center (Botany section), Darwinweg 2, 2333 CR Leiden, the Netherlands. ³⁶Biosystematics Group, Wageningen University, Droevendaalsesteeg 1, 6708 PB Wageningen, the Netherlands. ³⁷Department of Ecology, Faculty of Science, Charles University in Prague, CZ-128 44 Viničná 7, Prague 2, Czech Republic. ³⁸Centre for Invasion Biology, Department of Botany and Zoology, Stellenbosch University, Matieland 7602, South Africa.

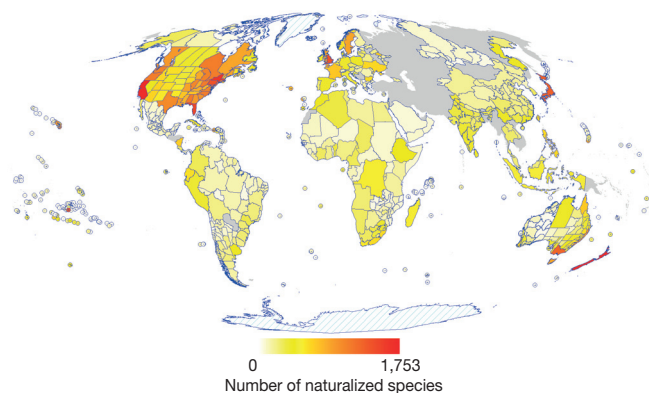


Figure 1 | Naturalized vascular plant species in the 843 regions covered by the GloNAF database. The heat-map colours correspond to the number of naturalized species in each of the regions (including 362 island regions). Areas permanently covered by ice sheets are indicated in hatched cyan. Grey areas indicate regions lacking naturalized plant data. To allow comparisons between the sizes of the GloNAF regions, we used a Mollweide equal-area projection. However, to increase the visibility of small islands and island groups on the map, they are represented by circles.

We found that at least 13,168 vascular plant species have become naturalized in at least one of the 843 regions (including 362 islands) (Fig. 1). As there were no data available for approximately 17% of the Earth's land area, particularly in temperate Asia (Fig. 1), and some of the regional inventories used might not be fully comprehensive, the actual number is likely to be even higher. This means that at least 3.9% of all currently known vascular plant species on Earth ($n = 337,137$; <http://www.theplantlist.org/>) have become naturalized outside their natural ranges as a result of human activity. With continuing globalization and increasing international traffic and trade, it is very likely that more species will be introduced outside their natural ranges and naturalize.

To assess which continents have accumulated the highest number of naturalized species, we assigned each of the GloNAF regions to the nine major biogeographically defined areas recognized by the Biodiversity Information Standards (also known as the Taxonomic Databases Working Group (TDWG)¹⁴; Fig. 2a). Since the areas of the TDWG continental scheme (further referred to as TDWG continents) differ significantly in size, we created accumulation curves of naturalized species to allow comparisons of the number of naturalized plants per continent for equal areas¹⁵. When ignoring differences in total area, North America has the highest cumulative number of naturalized species ($n = 5,958$), followed by Europe ($n = 4,140$; Fig. 2b). Although the rich naturalized floras of these continents could partly reflect a higher sampling intensity in these continents, it is likely that they also reflect a higher introduction effort. Both continents have dominated international trade for centuries, and many plants have been intentionally introduced from other continents for agricultural and horticultural purposes^{16,17}.

Although North America has a longer history of European colonization than Australasia, it received only slightly more naturalized species from outside the continent (3,513) than the latter (3,371; Fig. 2c). However, Australasia has even more such extra-continental species than North America when taking into account area differences (Fig. 2c). One possible explanation is that Australia's long biogeographical isolation and drying climate have resulted in a native flora that is phylogenetically distinct¹⁸, but not well-adapted to exploit the novel habitats created by European settlers. These new habitats have instead been occupied by many incoming alien plant species.

When only extra-continental arrivals are considered, Europe drops to fifth position, just behind Africa (Fig. 2c). Thus, although many plants from other continents have been introduced into Europe^{17,19}, few of them have naturalized. One explanation might be that plants

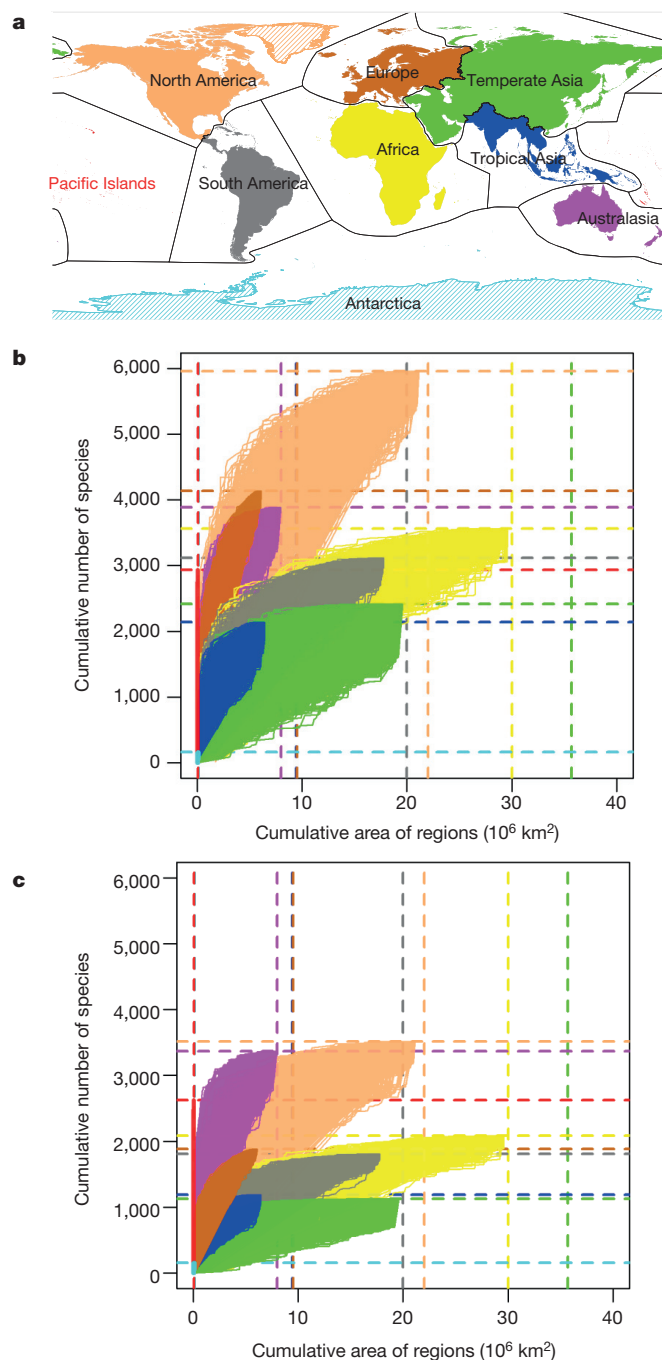


Figure 2 | Naturalized species-accumulation curves for the major biogeographical areas. **a**, Map of the nine TDWG continents. Hatched areas indicate major permanent ice sheets. **b**, Naturalized species-accumulation curves (1,000 random draws) for each of the nine continents. **c**, Same as **b** but here naturalized species are restricted to extra-continental aliens only. The colours in **b** and **c** correspond to the colours of the continents in **a**. Vertical and horizontal dashed lines mark the total area of the continent and its total number of naturalized plants, respectively. To increase visibility, thicker lines were used for Pacific Islands and Antarctica.

that spread through Europe with agriculture several thousand years ago (so-called archaeophytes), and European species that naturalized within the continent more recently, have already occupied many of the vacant niches, preventing many extra-continental species from naturalizing. In addition, extra-continental species might be relatively maladapted to the human-dominated environments in Europe, compared with species already present there, which have a longer evolutionary history of growing in these environments⁶.

The Pacific Islands show the steepest increase in cumulative number of naturalized species with area (Fig. 2). Therefore, our data provide the first global test illustrating that oceanic islands harbour more naturalized alien plants than similarly sized mainland regions, a phenomenon that is attributed to the available niche space not being saturated by native species^{8,20} or to a higher number of introductions. Given the high concentration of endemic species on most oceanic islands²¹, the great richness of naturalized species on these islands constitutes a serious threat to global biodiversity.

TDWG continents with large tropical regions (Africa, South America and tropical Asia) have, overall, fewer naturalized alien species than the predominantly temperate continents (North America, Europe and Australasia). This is consistent with previous observations suggesting a higher resistance of tropical regions to the establishment of alien species because of fewer available free ecological niches, faster recovery of vegetation after disturbance or a lower introduction rate^{7,22,23}. Temperate Asia, in contrast, shows a very low rate of accumulation of naturalized species with area. Unlike other continents, most of temperate Asia has not been colonized by Europeans (http://commons.wikimedia.org/wiki/Atlas_of_colonialism), and large parts of it have only recently opened up to inward movements of people and plants²⁴. With the recent rise of China as a major trade partner, we might expect a rapid increase of naturalized species in temperate Asia in the coming decades.

To identify the major donor continents of naturalized alien plant species, we assigned each naturalized species to its native continent(s). On the basis of estimated numbers of native species per continent, one would expect the most species-rich TDWG continents (South America and tropical Asia) to be the main donors of naturalized plant species (Fig. 3a); but they are not. The observed flow of naturalized plant species clearly shows that temperate Asia and Europe are the major donors (Fig. 3b). Although temperate Asia is ahead of Europe in absolute numbers, the observed number of species native to Europe and naturalized elsewhere is 288% higher than expected, but only 52% higher than expected for temperate Asia (Extended Data Fig. 1 and Extended Data Table 1). Furthermore, North America is also over-represented, with 57% more species donated than expected (Extended Data Fig. 1). In contrast, the TDWG continents that are largely in the Southern Hemisphere are all underrepresented as donors (Extended Data Fig. 1). These results are robust against potential over- or underestimates of the number of native species per continent (see Extended Data Table 1 for a sensitivity analysis). This suggests that the traditionally acknowledged Old World versus New World dichotomy in

biological invasions^{6,7} needs to be replaced by a Northern Hemisphere versus Southern Hemisphere dichotomy for the donor continents of naturalized alien plants globally. Darwin²⁵ suggested that Northern Hemisphere species, as a consequence of a more competitive evolutionary history, are intrinsically better competitors than Southern Hemisphere species, and that this could explain their naturalization success. To determine whether this is indeed the case requires further research. Nevertheless, the fact that the Southern Hemisphere is currently underrepresented as a donor might also indicate that the southern continents still harbour many species that could potentially spread to northern continents when given the chance.

For six of the nine TDWG continents, the observed intra-continental flows were larger than expected (Fig. 3 and Extended Data Fig. 1). Because of the shorter distances, intra-continental propagule pressure can be assumed to have been larger, and because of environmental similarity, subsequent naturalization chances are higher for intra-continental alien species²⁶. Notable exceptions with fewer than expected intra-continental naturalizations were South America and tropical Asia. We argue that because many species from these continents have restricted ranges—reflected in relatively high levels of regional endemism²¹—species from tropical Asia and South America are less likely to have been dispersed outside their native ranges.

The recently compiled GloNAF database has enabled the most comprehensive analysis so far of the global distributions of naturalized alien plant species, and provides the first robust estimates of the flows of naturalized plant species worldwide. We reveal striking differences within and among continents in the sizes of their naturalized alien floras, rates of accumulation of naturalized species with respect to area, and relative importance as exporters of naturalized species. Humans have strongly shaped the geographical composition and global distribution of alien plants among the World's continents, with the Northern Hemisphere being the major donor. The Pacific Islands and Australasia harbour the highest numbers of naturalized alien species, given their sizes and the extent of naturalization of species from other continents. The GloNAF database and the robust large-scale patterns we reveal here provide a vital foundation for testing fundamental hypotheses to understand plant naturalization better. For example, when combined with native plant inventories and phylogenetic data, the database will allow quantification of the degree of global floristic homogenization and tests to determine whether naturalized species are more closely or more distantly related to native species²⁵. In addition, the global baseline data of plant naturalizations provided here might contribute an essential biodiversity variable needed to monitor

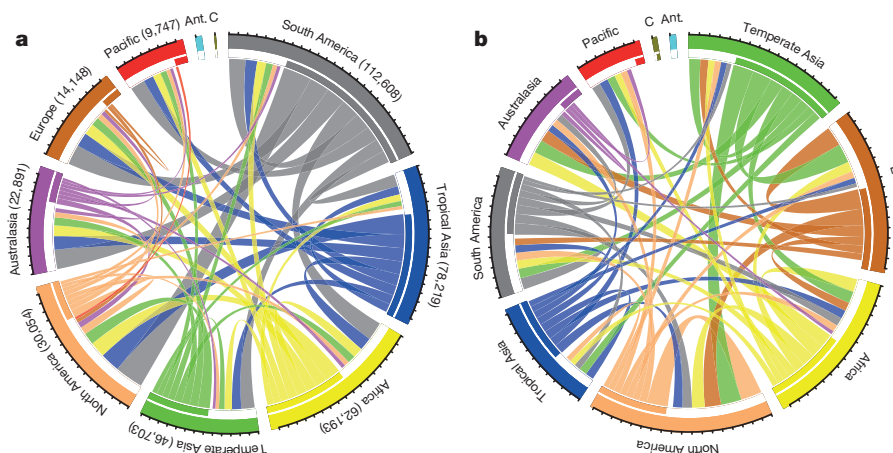


Figure 3 | Flows of naturalized alien plant species among the TDWG continents. **a**, Expected flows (medians of 999 random draws) of naturalized species on the basis of estimated numbers of native species (in brackets). **b**, Observed flows of naturalized species. The continents are ordered according to decreasing importance as sources. Only the 50% most important flows are

shown. Ant., Antarctica ($n = 293$ native species); C, only known from cultivation or novel hybrids ($n = 97$ species). Each tick along the outer circle corresponds to 1,000 species. Left (white) parts of inner bars along the circle represent flows of imported species; right (coloured) parts represent exported species.

changes in global biodiversity¹³, and can inform evidence-based management of alien species.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 8 April; accepted 14 July 2015.

Published online 19 August 2015.

- Steffen, W., Grinevald, J., Crutzen, P. & McNeill, J. The Anthropocene: conceptual and historical perspectives. *Phil. Trans. R. Soc. A* **369**, 842–867 (2011).
- Lewis, S. L. & Maslin, M. A. Defining the Anthropocene. *Nature* **171**, 171–180 (2015).
- Simberloff, D. *et al.* Impacts of biological invasions: what's what and the way forward. *Trends Ecol. Evol.* **28**, 58–66 (2013).
- Richardson, D. M. *et al.* Naturalization and invasion of alien plants: concepts and definitions. *Divers. Distrib.* **6**, 93–107 (2000).
- Blackburn, T. M. *et al.* A proposed unified framework for biological invasions. *Trends Ecol. Evol.* **26**, 333–339 (2011).
- di Castri, R. in *Biological Invasions: a Global Perspective* (eds Drake, J. A. *et al.*) 1–30 (John Wiley, 1989).
- Lonsdale, W. M. Global patterns of plant invasions and the concept of invasibility. *Ecology* **80**, 1522–1536 (1999).
- Elton, C. S. *The Ecology of Invasions by Animals and Plants* Ch. 4, 77–93 (Methuen, 1958).
- Blackburn, T. M., Cassey, P. & Lockwood, J. L. The island biogeography of exotic bird species. *Glob. Ecol. Biogeogr.* **17**, 246–251 (2008).
- Fridley, J. D. Of Asian forests and European fields: eastern U.S. plant invasions in a global floristic context. *PLoS ONE* **11**, e3630 (2008).
- Richardson, D. M. & Rejmánek, M. Trees and shrubs as invasive alien species – a global review. *Divers. Distrib.* **7**, 788–809 (2011).
- Lowe, S., Browne, M., Boudjelas, S. & De Poorter, M. *100 of the World's Worst Invasive Alien Species* (The Invasive Species Specialist Group (ISSG), 2000).
- Pereira, H. M. *et al.* Essential biodiversity variables. *Science* **339**, 277–278 (2013).
- Brummit, R. K. *World Geographical Scheme for Recording Plant Distributions Edition 2* (Hunt Institute for Botanical Documentation, 2001).
- Scheiner, S. M. Six types of species-area curves. *Glob. Ecol. Biogeogr.* **12**, 441–447 (2003).
- Brickell, C. & Cathey, H. M. *American Horticultural Society A–Z Encyclopedia of Garden Plants* (DK Publishing, 2004).
- Cullen, J., Knees, S. G. & Cubey, H. S. *The European Garden Flora* (Cambridge Univ. Press, 2011).
- Crisp, M., Cook, L. & Steane, D. Radiation of the Australian flora: what can comparisons of molecular phylogenies across multiple taxa tell us about the evolution of diversity in present-day communities? *Phil. Trans. R. Soc. Lond. B* **359**, 1551–1571 (2004).
- Heywood, V. H. & Sharroc, S. *European Code of Conduct for Botanic Gardens on Invasive Alien Species* (Council of Europe, 2013).
- Denslow, J. S. Weeds in paradise: thoughts on the invasibility of tropical islands. *Ann. Mo. Bot. Gard.* **90**, 119–127 (2003).
- Kier, G. *et al.* A global assessment of endemism and species richness across island and mainland regions. *Proc. Natl Acad. Sci. USA* **106**, 9322–9327 (2009).
- Rejmánek, M. in *Biodiversity and Ecosystem Processes in Tropical Forests* (eds Orians, G. H., Dirzo, R. & Cushman, J. H.) 153–172 (Springer-Kluwer, 1996).
- Fridley, J. D. *et al.* The invasion paradox: reconciling pattern and process in species invasions. *Ecology* **88**, 3–17 (2007).
- Keller, W., Li, B. & Shiue, C. H. China's foreign trade: perspectives from the past 150 years. *World Econ.* **34**, 853–892 (2011).
- Darwin, C. *On the Origin of Species by Means of Natural Selection, or the Preservation of Favoured Races in the Struggle for Life* (John Murray, 1859).
- Lambdon, P. W. *et al.* Alien flora of Europe: species diversity, temporal trends, geographical patterns and research needs. *Preslia* **80**, 101–149 (2008).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank the DAISIE team, the CONABIO team, E. Zykova, J. K. Vinogradova, S. R. Majorov, M. Schmidt, M. Newman, P. Thomas, R. Pooma, S. McCune, S. S. Tjitroedirdjo, H. Roy, S. Rorke, J. Danihelka, Z. Barina, A. Zeddard, S. Masciadri, Z. Barina and P. Nowak for data contributions, Z. Sixtová, B. Rüter, E. Mamonova, M. Krick, O. Michels and T. Scheu for digitizing data and internet searches, G. Müller and J. Moat for help with shapefiles, L. Cayuela for help with the R package Taxonstand, and T. Blackburn, A. Meyer and M. Rejmánek for comments on previous versions of the manuscript. M.v.K. and W.D. acknowledge funding by the Deutsche Forschungsgemeinschaft (KL 1866/9-1). F.E. acknowledges funding by the Austrian Climate and Energy Fund (project number KR11ACOK00355, SpecAdapt). J.P. and P.P. were supported by the Centre of Excellence PLADIAS (Czech Science Foundation project number 14-36079G) and long-term research development project RVO 67985939 (The Czech Academy of Sciences). P.P. acknowledges support by Praemium Academiae award from The Czech Academy of Sciences. M.W. and M.S. acknowledge funding from the Helmholtz Centre for Environmental Research (UFZ) and the German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig (DFG FZT 118). P.W. and H.K. acknowledge funding from the Deutsche Forschungsgemeinschaft (DFG) Free Floater Program in the Excellence Initiative at the University of Göttingen and in the scope of the BEFmate project from the Ministry of Science and Culture of Lower Saxony. H.S. acknowledges support by the German VW-Foundation. F.J.C. and M.V. acknowledge support of the project Flora de Guinea Ecuatorial, 4 (CGL2012-32934). N.F. thanks the Projects ICM 05-002, PFB-23 and Fondecyt Postdoc 3120125. J.T. acknowledges the support of the Research Center of the College of Science, King Saud University, Riyadh, Saudi Arabia.

Author Contributions M.v.K., P.P., W.D., F.E., J.P., E.W., M.W., H.K. and P.W. are the core GloNAF project members, who searched for and coordinated the collection of inventories of naturalized alien plants. M.v.K. and P.P. digitized the inventories and standardized the taxonomic names. J.K., N.M., L.A., J.B., F.C., D.C., J.C.-T., N.C., E.C., C.C., A.E., E.F., N.F., Q.G., L.H., I., A.K., S.M., J.M., O.M., D.M., D.N., A.P., P.P., M.P.B., M.P., M.S., H.S., W.S., J.T., M.V. and J.W. contributed naturalized plants inventories or other data. M.v.K. led the analyses and writing, with major inputs from P.P., W.D., F.E., J.P., M.W., H.K. and P.W., and further inputs from all other authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.v.K. (mark.vankleunen@uni-konstanz.de).

METHODS

Data compilation. No statistical methods were used to predetermine sample size. The GloNAF database includes inventories of naturalized alien plant species (also infraspecific taxa and hybrid taxa) for 843 regions worldwide. The data sources that we used (see Supplementary Information) include naturalized alien plant compendia, national and subnational lists of naturalized alien plant species published in scientific journals, as books or on the internet, as well as books and online compendia of national or subnational floras with information on which species occur in the wild but are not native. Our database also includes unpublished inventories of naturalized alien species that were specifically compiled for the GloNAF database (for example, for the provinces of China and the states of India). We consider those alien species that have established self-sustaining populations without direct human intervention to be naturalized, following refs 4 and 27. The GloNAF database will be fully publicly available after finalizing funded GloNAF projects (Deutsche Forschungsgemeinschaft and Austrian Science Fund FWF), which are due in about 3 years.

As certain regions of the world are more intensively researched than others, it is unavoidable that some of the regional inventories of naturalized alien species are more comprehensive than others. We aimed to include the most comprehensive and most recent regional inventories. Indeed, more than 95% of the data sources are from the past two decades (see Supplementary Data). Moreover, since some of the original source lists included alien species that are cultivated only or have non-persistent populations in the wild, we excluded those species whenever such information was provided, or contacted experts of the regional floras to remove species of doubtful naturalization status. Furthermore, for European countries that differentiated between archaeophytes (alien species that came before the year 1492) and neophytes (species that came after the year 1492), we kept only the latter, because the alien status of some species classified as archaeophytes is disputed; moreover, this classification is not available for other regions of the world, and thus would prevent us from achieving a balanced/standardized assessment of naturalized alien species numbers.

To standardize scientific names, each naturalized plant inventory was compared with The Plant List (<http://www.theplantlist.org/>), the most comprehensive working list of all plant species²⁸. This taxonomic standardization was done with the help of the R²⁹ package Taxonstand³⁰. For each species, we kept the name accepted by The Plant List. Species that were not found in The Plant List, also not after accounting for spelling differences, were kept in the database using the names as used in the source data. In total, the database includes 13,168 species of which 13,033 are recognized by The Plant List (12,498 as accepted and 535 as unresolved names). The remaining 135 species do not occur in The Plant List, and among those 11 are ornamental cultivars.

For each species in the database, we compiled data on which of the nine regions of the TDWG continental scheme (TDWG continents¹⁴) it was native to, or whether it was known only from cultivation or resulted from hybridization between two alien species or an alien and a native species. Most of the native-range data were extracted from the World Checklist of Selected Plant Families (WCSP; <http://apps.kew.org/wcsp/>), and supplemented with data from the Germplasm Resources Information Network (<http://www.ars-grin.gov/cgi-bin/npgs/html/index.pl>). For the approximately 4,000 species that were not included in these two major data sources, we retrieved information on the native regions from printed floristic compendia, extensive internet searches and comparisons of their naturalized distributions to their overall distributions in the Global Biodiversity Information Facility (<http://www.gbif.org/>). Information about native continents was found for 13,070 species, of which 219 are only known from cultivation and 51 are novel hybrids. Many (5,646) species were native to more than one continent. For the few (98) remaining species, we could not find any information on their native ranges.

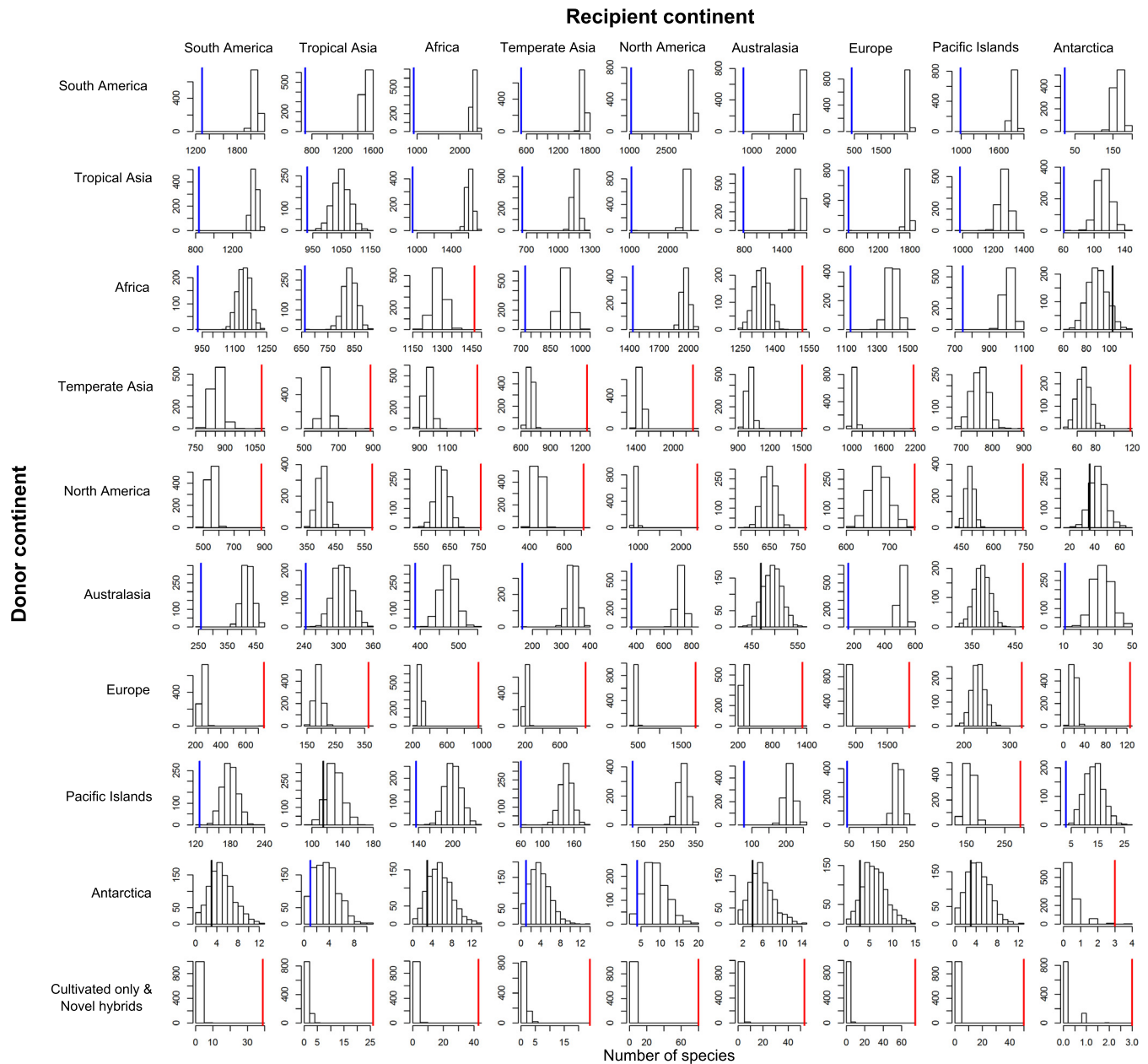
Each of the 843 regions covered by GloNAF was assigned to one of the nine TDWG continents. We calculated the area of each region while considering only the ice-sheet-free areas of each region, ranging from 0.03 to 2,486,952 km², with a median of 18,725 km².

Accumulation of naturalized species per continent. To determine which continent accumulated the highest number of naturalized species for a certain area, we constructed species-accumulation curves¹⁵ separately for each of the nine TDWG continents. Since choosing a starting region and the order of adding remaining regions to the species-accumulation curves would be arbitrary, we used a random order of regions, and repeated this procedure 1,000 times. Species-accumulation curves were calculated for all alien species and for extra-continental alien species separately. This analysis was done in the R package *vegan*³¹.

Flows of naturalized alien species among continents. To test whether the observed flows of naturalized species from donor continents to recipient continents were larger or smaller than expected, we compared the observed flows with those based on random draws from the extant global flora. Since no data on the number of native species per TDWG continent exist, we first estimated these numbers by extrapolation of the known native origins of 130,641 accepted vascular plant species in the WCSP (<http://apps.kew.org/wcsp/>) to the total number of 337,137 accepted species in The Plant List (<http://www.theplantlist.org/>). Although the WCSP includes quite a large proportion (38.8%) of all vascular plant species, it does not include all vascular plant families yet, and it might be geographically biased. However, ref. 32 showed that all 52 TDWG level-2 regions, and thus the TDWG continents also, are well represented in the WCSP. Furthermore, our estimates did not deviate much from published estimates we found for some of the continents: our estimate of 62,193 native species for Africa is close to the previously estimated 40,000–60,000 for the African mainland³³, and the 64,500 species listed in the African Plants Database (<http://www.ville-ge.ch/musinfo/bd/cjb/africa/>). Our estimate of 14,148 native species for Europe is slightly higher than the 12,517 native species listed in the Flora Europaea³⁴. Our estimate of 30,054 native species for North America is higher than the 21,500 species listed in the Biota of North America Program (<http://www.bonap.org/>), but the latter does not include species of Mexico. Our estimate of 22,891 native species for Australasia is higher than the 19,324 reported for Australia by the Australian National Herbarium (<https://www.anbg.gov.au/aust-veg/australian-flora-statistics.html>), but the latter does not cover all parts of Australasia (for example, New Zealand). Therefore, although our estimates of the native species richness of each continent are higher than previous estimates, these differences seem to result mainly from additional regions included in TDWG continents and gaps in the other data sources. Thus our results appear to be realistic proxies for the true numbers of continental species richness.

To obtain the expected flows of species from donor to recipient continents, we first created a species pool with a size equal to that of the extant global vascular-plant species pool ($n = 337,137$), in which the proportion of species native to each continent or combination of continents was based on the estimated native species richness of the continents. Then, for each recipient continent, we drew separately a random sample of species from the extrapolated global species pool. The size of the random sample was equal to the number of naturalized alien species observed in the recipient continent. We then recorded the number of randomly drawn species native to each continent or belonging to the pool of species known from cultivation or as novel hybrids. This random-draw procedure was repeated 999 times, and the medians are shown in Fig. 3a. We did this for each recipient continent separately to allow for the fact that a species can be naturalized in more than one continent. If the observed flow of species from a donor continent to a recipient continent was within the upper 2.5% of the random distribution, we considered the observed flow to be significantly larger than expected by chance; if the observed flow was within the lower 2.5% of this distribution, we considered the flow to be significantly lower than expected by chance. Since we might have over- or underestimated the native species richness for some continents, we also did a sensitivity analysis by decreasing and increasing the size of the native flora of each continent by 10% in turn (see Supplementary Information). R syntax for the random draws is available from the corresponding author on request. Flow plots were created using an R syntax adapted from ref. 35.

27. Pyšek, P. *et al.* Alien plants in checklists and floras: towards better communication between taxonomists and ecologists. *Taxon* **53**, 131–143 (2004).
28. Kalwij, J. M. Review of 'The Plant List, a working list of all plant species'. *J. Veg. Sci.* **23**, 998–1002 (2012).
29. R Core Team. R: a language and environment for statistical computing v.3.1.2 (R Foundation for Statistical Computing, 2014).
30. Cayuela, L. & Oksanen, J. Taxonstand: taxonomic standardization of plant species names v.1.6 (R Foundation for Statistical Computing, 2014).
31. Oksanen, J. *et al.* *Vegan: community ecology package v.2.0-10* (R Foundation for Statistical Computing, 2013).
32. Joppa, L., Visconti, P., Jenkins, C. N. & Pimm, S. L. Achieving the convention on biological diversity's goals for plant conservation. *Science* **341**, 1100–1103 (2013).
33. Scholes, R. J. *et al.* in *Africa Environment Outlook 2* Ch. 7, 226–261 (African Ministerial Conference on the Environment and United Nations Environment Programme, 2006).
34. Tutin, T. G. *et al.* *Flora Europaea* (Cambridge Univ. Press, 1964–1980).
35. Abel, G. J. & Sander, N. Quantifying global international migration flows. *Science* **343**, 1520–1522 (2014).



Extended Data Figure 1 | Observed and expected numbers of naturalized species from each donor TDWG continent in each of the recipient TDWG continents. Histograms of the expected numbers are shown in black open bars, and are based on 999 random draws from the global flora ($n = 337,137$). The observed numbers are shown as vertical lines; blue, significantly fewer

observed naturalized species from the source continent than expected (in the lower 2.5 percentile); red, significantly more observed naturalized species than expected (in the upper 2.5 percentile); black, the observed number of naturalized species is within the central 95% range of the expected numbers.

Extended Data Table 1 | Results of sensitivity analysis for observed and expected numbers of naturalized species from each donor continent in each of the recipient continents

		Recipient continent																	
		South America		Tropical Asia		Africa		Temperate Asia		North America		Australasia		Europe		Pacific Islands		Antarctica	
		No.	P	No.	P	No.	P	No.	P	No.	P	No.	P	No.	P	No.	P	No.	P
Donor continent	South America	1293		714		931		501		1070		763		399		993		24	
		2071	0.000	1509	0.000	2323	0.000	1676	0.000	3568	0.000	2432	0.000	2536	0.000	1834	0.000	163	0.000
		1927	0.000	1405	0.000	2166	0.000	1563	0.000	3317	0.000	2263	0.000	2361	0.000	1707	0.000	151	0.000
	Tropical Asia	2205	0.000	1608	0.000	2474	0.000	1785	0.000	3791	0.000	2588	0.000	2698	0.000	1954	0.000	173	0.000
		837		931		950		662		1054		787		641		983		61	
		1438	0.000	1050	0.000	1613	0.000	1165	0.000	2473	0.000	1687	0.000	1759	0.000	1272	0.000	113	0.000
	Africa	1325	0.000	967	0.000	1490	0.000	1074	0.000	2283	0.000	1555	0.000	1624	0.000	1175	0.000	104	0.000
		1547	0.000	1128	0.095	1738	0.000	1255	0.000	2662	0.000	1815	0.000	1891	0.000	1370	0.000	122	0.000
	Temperate Asia	930		663		1464		721		1434		1521		1127		747		103	
		1144	0.000	834	0.000	1285	0.999	924	0.000	1968	0.000	1342	0.999	1401	0.000	1013	0.000	90	0.928
		1047	0.000	764	0.000	1175	0.983	849	0.000	1806	0.000	1233	0.982	1285	0.000	931	0.000	83	0.734
	North America	1235	0.000	901	0.000	1387	0.999	1002	0.000	2126	0.000	1451	0.999	1512	0.000	1095	0.000	97	0.991
		1085		886		1320		1272		2315		1505		2166		893		118	
		860	0.999	626	0.999	964	0.999	696	0.999	1478	0.999	1010	0.999	1049	0.999	761	0.999	67	0.999
	Australasia	784	0.999	570	0.999	879	0.999	632	0.999	1348	0.999	920	0.999	961	0.999	695	0.994	62	0.999
		931	0.999	680	0.999	1047	0.999	755	0.999	1604	0.999	1095	0.999	1142	0.999	826	0.999	73	0.999
	Europe	880		577		758		714		2376		775		759		737		36	
		553	0.999	403	0.999	621	0.999	448	0.999	951	0.999	648	0.999	677	0.999	490	0.999	43	0.111
		502	0.999	367	0.999	562	0.998	407	0.999	863	0.999	589	0.993	613	0.811	444	0.999	39	0.026
	Pacific Islands	603	0.999	439	0.999	675	0.999	487	0.999	1037	0.999	707	0.999	737	0.999	534	0.999	47	0.252
	Antarctica	259		243		387		165		364		470		163		468		11	
		421	0.000	307	0.000	474	0.000	340	0.000	724	0.000	496	0.105	516	0.000	374	0.999	33	0.000
		381	0.000	278	0.000	429	0.000	309	0.000	656	0.000	449	0.001	467	0.000	337	0.998	30	0.000
	Cultivated only & novel hybrids	460	0.000	334	0.014	514	0.011	371	0.000	793	0.000	540	0.863	562	0.000	407	0.999	36	0.000
		746		364		966		901		1839		1322		2213		326		126	
		261	0.999	190	0.999	292	0.999	211	0.999	448	0.999	305	0.999	319	0.999	231	0.999	20	0.999
	Europe	234	0.999	172	0.999	264	0.999	190	0.999	404	0.999	276	0.999	289	0.999	208	0.999	18	0.999
		285	0.999	207	0.999	321	0.999	231	0.999	489	0.999	335	0.999	335	0.999	253	0.999	22	0.999
	Pacific Islands	127		115		136		60		131		78		43		291		3	
		180	0.000	130	0.069	201	0.000	146	0.000	308	0.000	210	0.000	220	0.000	158	0.999	14	0.000
		162	0.000	117	0.002	181	0.000	131	0.000	278	0.000	190	0.000	198	0.000	143	0.999	13	0.000
	Antarctica	197	0.003	144	0.398	221	0.000	159	0.000	338	0.000	232	0.000	241	0.000	174	0.999	15	0.001
		3		1		3		1		4		4		3		3		3	
		5	0.093	4	0.022	6	0.059	4	0.013	9	0.015	6	0.118	6	0.045	5	0.136	0	0.992
	Cultivated only & novel hybrids	5	0.071	3	0.014	5	0.040	4	0.008	8	0.007	6	0.092	6	0.027	4	0.105	0	0.985
		6	0.144	4	0.027	6	0.083	5	0.018	10	0.028	7	0.191	7	0.071	5	0.183	0	0.996
	Europe	39		26		43		24		80		53		75		50		3	
		2	0.999	1	0.999	2	0.999	1	0.999	3	0.999	2	0.999	2	0.999	1	0.999	0	0.997
		1	0.999	1	0.999	2	0.999	1	0.999	3	0.999	2	0.999	2	0.999	1	0.999	0	0.996
	Cultivated only & novel hybrids	2	0.999	1	0.999	2	0.999	1	0.999	3	0.999	2	0.999	2	0.999	2	0.999	0	0.999

For each combination of two TDWG continents, the table gives the observed number (No.) of species that are native to the donor continent and have become naturalized in the recipient continent in bold type. Below, each observed number is the median of the expected number on the basis of 999 random draws from the global vascular flora ($n = 337,137$). Below this median, the minimum and maximum median values of the expected numbers found during the sensitivity analysis are given in sloping type. In addition, the table gives the proportion (P) of the 999 random draws for the expected values that were smaller than the observed values. The minimum and maximum proportions found during the sensitivity analysis are given in sloping type. Proportions >0.975 (the source is overrepresented in the recipient continent) are given in red; proportions <0.025 (the source is underrepresented) are given in blue.

Genetic evidence for two founding populations of the Americas

Pontus Skoglund^{1,2}, Swapan Mallick^{1,2,3}, Maria Cátira Bortolini⁴, Niru Chennagiri^{1,2}, Tábita Hünemeier⁵, Maria Luiza Petzl-Erler⁶, Francisco Mauro Salzano⁴, Nick Patterson² & David Reich^{1,2,3}

Genetic studies have consistently indicated a single common origin of Native American groups from Central and South America^{1–4}. However, some morphological studies have suggested a more complex picture, whereby the northeast Asian affinities of present-day Native Americans contrast with a distinctive morphology seen in some of the earliest American skeletons, which share traits with present-day Australasians (indigenous groups in Australia, Melanesia, and island Southeast Asia)^{5–8}. Here we analyse genome-wide data to show that some Amazonian Native Americans descend partly from a Native American founding population that carried ancestry more closely related to indigenous Australians, New Guineans and Andaman Islanders than to any present-day Eurasians or Native Americans. This signature is not present to the same extent, or at all, in present-day Northern and Central Americans or in a ~12,600-year-old Clovis-associated genome, suggesting a more diverse set of founding populations of the Americas than previously accepted.

All Native American groups studied to date can trace all or much of their ancestry to a single ancestral population that probably migrated across the Bering land bridge from Asia more than 15,000 years ago², with some Northern American and Arctic groups also tracing other parts of their ancestry to more recent waves of migration^{2,9,10}. Ancient genomic evidence has shown that this so-called ‘First American’ ancestry is present in an individual associated with Clovis technology from North America dating to ~12,600 years ago³, and mitochondrial DNA has suggested that it was also present by 13,000–14,500 years ago^{11,12}. In contrast, some morphological analyses of early skeletons in the Americas have suggested that characteristics of some Pleistocene and early Holocene skeletons fall outside the variation of present-day Native Americans and instead fall within the variation of present-day indigenous Australians, Melanesians and so-called ‘Negrito’ groups from Southeast Asia (and some sub-Saharan African groups)^{7,13}. This morphology has been hypothesized to reflect an initial ‘Paleoamerican’ pioneer population in the Americas, which according to some interpretations was largely replaced by populations with Northeast Asian affinities in the early Holocene, but may have persisted in some locations^{14,15}. However, morphological similarity can arise not only through shared descent but also through convergent evolution or phenotypic plasticity coupled with similar environments^{16,17}. Another limitation of morphological data is that it provides very few independent characters that can be analysed. Genome-wide data, with its hundreds of thousands of independent characters that evolve effectively neutrally, should be a statistically powerful and robust way to test whether a distinct lineage contributed to Native Americans.

Analysis of population history in the Americas is complicated by post-Columbian admixture from mainly European and African sources². We identified 63 individuals without discernable evidence of European or African ancestry in 21 Native American populations genotyped at ~600,000 single nucleotide polymorphisms (SNPs) on

the Affymetrix Human Origins array^{18,19} (Extended Data Fig. 1 and Supplementary Information section 1). We further restricted our studies to individuals from Central and South America that have the strongest evidence of deriving entirely from a homogeneous First American ancestral population². We computed all possible f_4 -statistics of the form $f_4(\text{American}_1, \text{American}_2; \text{outgroup}_1, \text{outgroup}_2)$, the product of the allele frequency differences between the two American groups and the two outgroups. We represented the Americans by a panel of 7 Central and South American groups, and the outgroups by 24 populations (4 from each of 6 worldwide regions). If the two Native American groups descend from a homogeneous ancestral population whose ancestors separated from the outgroups at earlier times, it follows that the difference in allele frequencies between Native American populations will have developed entirely after their separation from the outgroups, and so the correlation in allele frequency differences is expected to be zero. To evaluate whether all possible f_4 -statistics computed in this way are consistent with zero, correcting for multiple hypothesis testing due to the large number of statistics examined, we measured the empirical covariance of the matrix of f_4 -statistics using a block jackknife¹⁸, and performed a single Hotelling’s T^2 test² for consistency with zero. We reject the null hypothesis at high significance ($P = 2 \times 10^{-7}$), suggesting that the analysed Native American populations do not all descend from a homogeneous ancestral population since separation from the outgroups (Extended Data Table 1 and Supplementary Information section 2). The coefficients for which non-American populations contribute the most to the signals separate Native Americans into a cline with two Amazonian groups (Suruí and Karitiana) on one extreme and Mesoamericans on the other (Extended Data Fig. 2). Among the outgroups, the most similar coefficients to Amazonian groups are found in Australasian populations: the Onge from the Andaman Islands in the Bay of Bengal (a so-called ‘Negrito’ group), New Guineans, Papuans and indigenous Australians (Supplementary Information section 2).

We extended our analysis to 197 non-American populations sampled worldwide^{18–20}. We computed D -statistics²¹ to test whether a randomly drawn derived allele from each worldwide population has an equal probability of matching a randomly drawn Mesoamerican or Amazonian chromosome at sites where these differ. This test takes as its null hypothesis the tree-like population history (*Test population*, (*Mesoamericans*, *Amazonians*)), and produces a positive D -statistic only in the case of excess affinity between the test population and Amazonians (negative values in the case of an excess affinity with Mesoamericans). Consistent with the signals observed when many populations are analysed together, we find that Andamanese Onge, Papuans, New Guineans, indigenous Australians and Mamanwa Negritos from the Philippines all share significantly more derived alleles with the Amazonians ($4.6 > Z > 3.0$ standard errors (s.e.) from zero) (Extended Data Table 2). No population shares significantly more derived alleles with the Mesoamericans than with the Amazonians. We

¹Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA. ²Broad Institute of Harvard and MIT, Cambridge, Massachusetts 02142, USA. ³Howard Hughes Medical Institute, Harvard Medical School, Boston, Massachusetts 02115, USA. ⁴Departamento de Genética, Instituto de Biociências, Universidade Federal do Rio Grande do Sul, 91501-970 Porto Alegre, RS, Brazil.

⁵Departamento de Genética e Biologia Evolutiva, Universidade de São Paulo, 05508-090, SP, Brazil. ⁶Departamento de Genética, Universidade Federal do Paraná, 81531-980 Curitiba, PR, Brazil.

find consistent results for this test not only for Onge, Papuans, New Guineans and indigenous Australians as representatives of Australasian populations, but also for different outgroups in place of chimpanzee: Africans, Europeans and East Asians ($2.8 < Z < 4.8$) (Supplementary Information section 3). In Fig. 1, we show a quantile–quantile plot of D -statistics contrasting the Mesoamerican Mixe and the Amazonian Suruí, revealing Australasian populations as the only discernible outliers.

We replicated the significant evidence for affinity between Australasians and Amazonians using D -statistics computed on Illumina SNP array data² (as an alternative to the Affymetrix Human Origins SNP array data) ($2.6 < Z < 3.0$) and on high-coverage genome sequences from 3 Yoruba, 2 Suruí, 3 Mixe and 16 Papuans (18 of these genomes are reported for the first time here^{22,23}; Table 1) ($Z = 4.3$). In addition to the three independent molecular experiments that these data sets represent, we find consistent results for all different mutation classes in the high-coverage genomes ($2.6 < Z < 4.3$), and different ascertainment schemes (for example, in polymorphisms discovered in Africans, New Guineans and East Asians) (Supplementary Information section 3) ($1.1 < Z < 3.3$ for panels with $>20,000$ SNPs). We also find consistent results for two differently genotyped subsets of Suruí individuals from a total of 24 individuals² (Table 1 and Extended Data Fig. 3a) ($2.6 < Z < 3.6$). Simulations (Supplementary Information

section 3) show that genotype and sequence errors cannot explain the magnitude of the observed signal (Extended Data Fig. 3b). Finally, we generated new data from 9 populations from present-day Brazil using the Affymetrix Human Origins array, including previously untested individuals from the Amazonian Suruí and Karitiana for which DNA was extracted from blood. These new samples replicate the signal, and furthermore show that the signal is also strong in the Xavante ($1.3 < Z < 3.25$), a population of the Brazilian Central Plateau which speaks a language of the Ge group that is different from the Tupi language group to which the languages of the Karitiana and Suruí both belong. We do not detect any excess affinity to Australasians in the $\sim 12,600$ -year-old Clovis-associated Anzick individual from western Montana ($Z = -0.6$) (Supplementary Information section 3).

To test if the significant D -statistics have the patterns expected for a genuine admixture event, we stratified the high coverage genomes into deciles of ‘ B -values’²⁴, which measures proximity to functionally important regions. Genuinely significant D -statistics are expected to be of larger magnitude closer to genes, as selection increases variability in fitness of haplotypes near functionally important regions, which in turn increases the genetic drift in these regions and the absolute magnitude of D -statistics^{25,26}, a prediction that we confirmed empirically (Extended Data Fig. 4). We computed D (Yoruba, Papuan; Mixe,

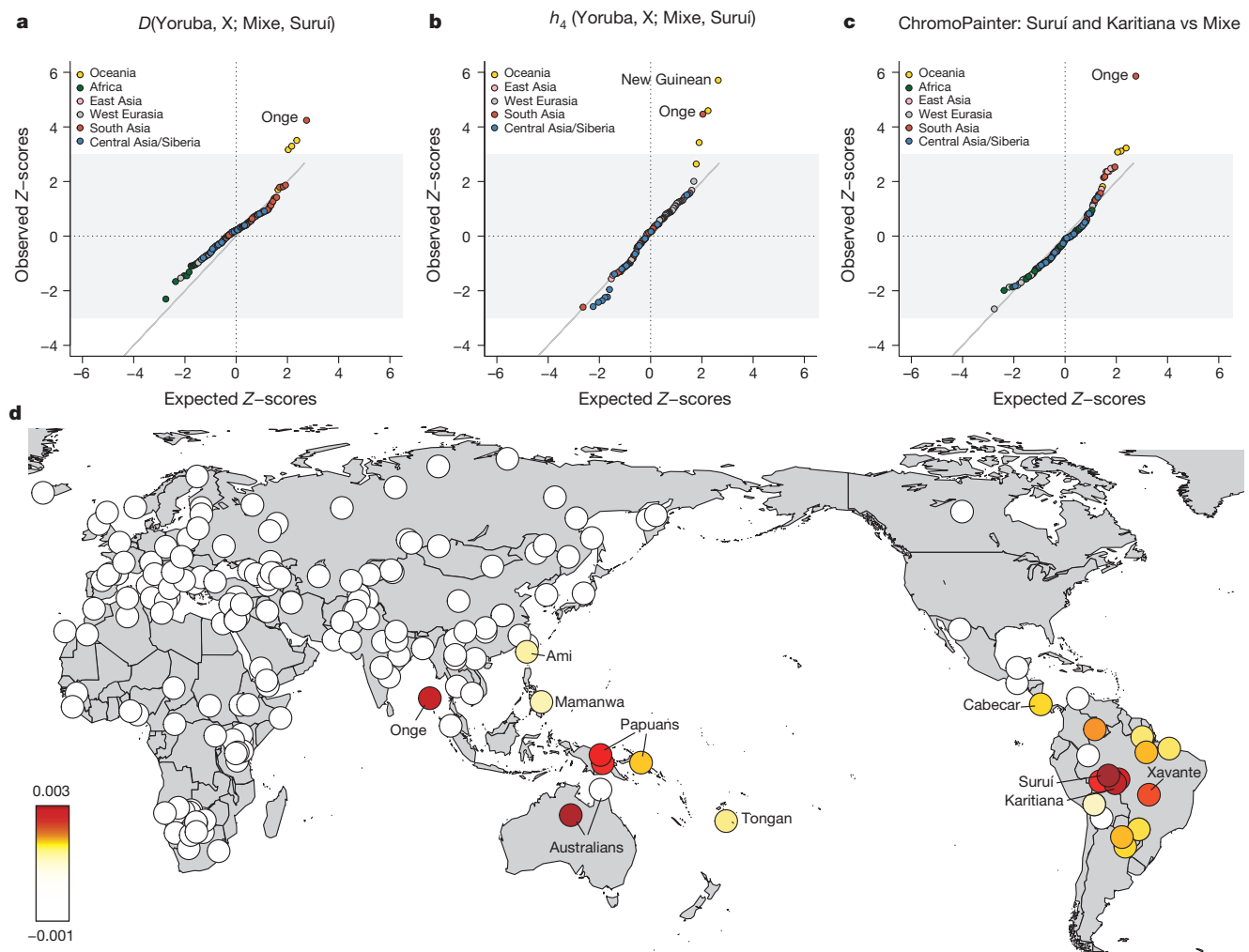


Figure 1 | South Americans share ancestry with Australasian populations that is not seen in Mesoamericans or North Americans. **a**, Quantile–quantile plot of the Z-scores for the D -statistic symmetry test for whether Mixe and Suruí share an equal rate of derived alleles with a candidate non-American population, X , compared to the expected ranked quantiles for the same number of normally distributed values. **b**, Z-scores for the h_4 -statistic. **c**, Z-scores for

the ChromoPainter statistic. **d**, Heatmap of ChromoPainter statistics. For non-Americans we display the symmetry statistic $S(\text{non-American}; \text{Mixe}, \text{Suruí and Karitiana})$ for donating as many haplotypes to Mixe as to Suruí and Karitiana. For the Americas we plot $S(\text{Onge}; \text{Mixe}, \text{American})$ for receiving as many haplotypes from the Onge as do the Mixe.

Table 1 | Statistics testing the consistency of the tree (Yoruba, (Papuan, (Mixe, Suruí))) with the data

	Test statistic	Z-score	Informative loci
High-coverage genomes	0.0211	4.26	798,873
A/T SNPs	0.0169	2.63	60,538
A/G SNPs	0.0191	3.64	268,962
A/C SNPs	0.0208	3.49	67,210
G/T SNPs	0.0248	4.27	67,623
C/T SNPs	0.0220	4.24	270,133
C/G SNPs	0.0248	4.26	64,951
Illumina array Suruí samples from HGDP	0.0076	2.63	247,814
Illumina array Suruí samples not in HGDP	0.0081	3.02	249,941
Affymetrix Human Origins array (Suruí cell lines)	0.0099	3.63	318,544
Affymetrix Human Origins array (Suruí blood samples)	0.0072	2.57	313,349
h_4 -statistic (Affymetrix Yoruba ascertainment)	0.0003	4.60	14,938
Chromosome painting symmetry test	0.0026	5.26	-

Note: except for the new h_4 statistics and chromosome painting symmetry tests which are explicitly noted, all statistics are D -statistics²¹. Z-scores were obtained by computing standard errors using a weighted block jackknife.

Suruí) separately for each bin, and found that it is of larger magnitude close to functionally important regions (Extended Data Fig. 4) ($Z = -2.0$ for the slope of a linear regression model), as expected for a real admixture event. A caveat is that when we formally combine the evidence from the genome-wide D -statistic and the correlation to the B -value, the significance ($Z = 3.6$ s.e. from 0) is not any greater than for the basic $D = 0.021 \pm 0.005$ statistic ($Z = 4.2$ s.e. from 0) because the two statistics co-vary. Nevertheless, the fact that the correlation with B -values is significant by itself and in the expected direction adds to the qualitative evidence for an admixture event.

Alternative approaches for testing for admixture involve detecting admixture linkage disequilibrium in a test population that is correlated to allele frequency differentiation between two populations that are related to the sources^{27,28}. We devised a statistic ' h_4 ' that is analogous to an f_4 -statistic, but instead of studying allele frequencies, it tests whether the linkage disequilibrium patterns of two populations are consistent with descending from a common ancestral population since separation from two outgroups. A classic statistic for measuring linkage disequilibrium in a population A is $H^A = p_{12}^A - p_1^A p_2^A$, which measures the extent to which a haplotype of two derived mutations occurring at frequency p_{12}^A is observed more or less frequently than would be expected from the individual frequencies of alleles 1 and 2 (p_1^A and p_2^A). Thus, we define $h_4(A, B; C, D)$ as the average of $(H^A - H^B)(H^C - H^D)$ across the genome, and view a deviation from zero as evidence against the unrooted tree $((A, B), (C, D))$. We used loci ascertained as polymorphic in African Yoruba, which is effectively an outgroup to the other populations analysed here, to test $h_4(\text{Yoruba}, X; \text{Mixe}, \text{Suruí})$ for all SNP pairs within 0.01 centimorgans (cM) and for a large set of worldwide non-African populations, and obtained normalized Z-scores by estimating the number of standard errors this quantity is from zero using a block jackknife. Although Z-scores computed for most of 120 non-Americans and non-Africans as population X conform to a normal distribution (Fig. 1b), we again found significant evidence of excess affinity of the Suruí to Australasian populations ($Z = 5.7$, $P = 10^{-8}$ for New Guineans; $Z = 4.6$, $P = 10^{-5}$ for Papuans; $Z = 4.4$, $P = 10^{-5}$ for Andamanese). When we exclude the Australasians, we detect no evidence of correlation between Z-transformed h_4 - and f_4 -statistics for the remaining 114 populations ($R = -0.026$) suggesting that h_4 can provide evidence independent of allele frequency based statistics. Although h_4 can theoretically be biased by loss of polymorphism due to bottlenecks (Supplementary Information section 4), there is no evidence that this is a problem for our analysis as East Asian and Siberian populations with comparable loss of polymorphism do not show an affinity to Amazonians by this statistic (Extended Data Fig. 5). In addition, there is a high degree of correlation between significant h_4 - and D -statistics in empirical data (Extended Data Fig. 5). Computing $h_4(\text{Yoruba}, \text{Onge}; \text{Mixe}, \text{Suruí})$ over windows of increasingly large genetic distances reveals that it dissipates at approximately 0.2 cM. This is an order of magnitude smaller than linkage disequilibrium caused by admixture events at

the ~4,000 year upper limit of previous methods¹⁸, but at a larger scale than the signal of admixture between Neanderthals and non-Africans 37,000–86,000 years ago²⁹ (Extended Data Fig. 5).

As a third population symmetry test, we applied a method for detecting shared haplotypes between individuals ('chromosome painting'³⁰) to infer in each Native American individual which non-American chromosome segment each American chromosome segment shares the closest affinity to, using a set of 174 non-American populations as references. We then performed a symmetry test for a candidate population sharing more haplotypes with a given non-American population than the Mesoamerican Mixe do, performing a block jackknife across all chromosomes (weighting to correct for variation in chromosome length) to assess uncertainty. We find that the blood and cell line Suruí are significantly closer to the Onge than the Mixe are ($Z = 5.3$) (Fig. 1c), as are the blood and cell line Karitiana samples ($Z = 4.2$ to 5.0), the Xavante ($Z = 4.3$), and the Piapoco and Guarani ($Z > 3$) (Fig. 1d). In contrast, populations from west of the Andes or north of the Panama isthmus show no significant evidence of an affinity to the Onge ($Z < 2$). An exception to this is the Cabecar, who have previously been shown to be partially admixed from a source south of the Panama isthmus².

The geographic distribution of the shared genetic signal between South Americans and Australasians cannot be explained by post-Columbian African, European or Polynesian gene flow into Native American populations. If such gene flow produced signals strong enough to affect our statistics, our statistics would show their strongest deviations from zero for African, European or Polynesian populations, which is not observed. For example, a direct test is significant in showing that the Suruí-specific ancestry component is genetically closer to the Andamanese Onge than to Tongans from Polynesia ($D = 0.0094$, $Z = 3.4$).

To investigate models consistent with the data, we studied admixture graph models relating the ancestry of Native American groups to Han Chinese and Onge Andaman Islanders, incorporating a previously described admixture event into Native American ancestors from a lineage related to an ~24,000-year-old Upper Paleolithic individual from Mal'ta in Siberia⁴ (denoted as MA1). We are unable to fit Amazonians as forming a clade with the Mesoamericans, or as having a different proportion of ancestry related to Mal'ta or present-day East Asians. Thus, our signal cannot be explained by lineages that have previously been documented as having contributed to Native American populations. However, we do find that a model where Amazonians receive ancestry from the lineage leading to the Andamanese fits the data in the sense that the predicted f_4 -statistics are all within two standard errors of statistics computed on the empirical data (Extended Data Figs 6 and 7 and Extended Data Table 3). These results do not imply that an unmixed population related anciently to Australasians migrated to the Americas. Although this is a formal possibility, an alternative model that we view as more plausible is that the 'Population Y' (after *Ypykuéra*, which means 'ancestor' in the Tupi language family spoken by the Suruí and

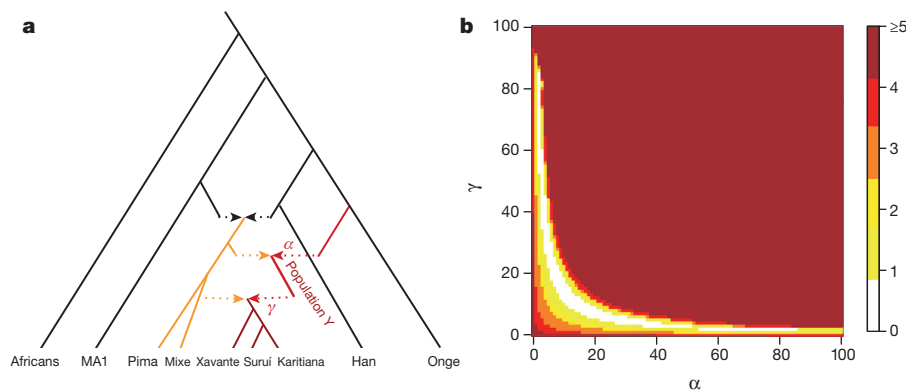


Figure 2 | A model of population history that can explain the excess affinity to Oceanians observed in Amazonian populations. **a**, We fit an admixture graph model where a population related to the Andamanese Onge contributed a fraction α of the ancestry of 'Population Y', which later contributed a fraction γ to the ancestry of Amazonian groups today

Karitiana) that contributed Australasian-related ancestry to Amazonians was already mixed with a lineage related to First Americans at the time it reached Amazonia. When we model such a scenario, we obtain a fit for models that specify 2–85% of the ancestry of the Suruí, Karitiana and Xavante as coming from Population Y (Fig. 2). These results show that quite a high fraction of Amazonian ancestry today might be derived from Population Y. At the same time, the results constrain the fraction of Amazonian ancestry that comes from an Australasian related population (via Population Y) to a much tighter range of 1–2% (Fig. 2).

We have shown that a Population Y that had ancestry from a lineage more closely related to present-day Australasians than to present-day East Asians and Siberians, likely contributed to the DNA of Native Americans from Amazonia and the Central Brazilian Plateau. This discovery is striking in light of interpretations of the morphology of some early Native American skeletons, which some authors have suggested have affinities to Australasian groups. The largest number of skeletons that have been described as having this craniofacial morphology and that date to younger than 10,000 years old have been found in Brazil⁶, the home of the Suruí, Karitiana and Xavante groups who show the strongest affinity to Australasians in genetic data. However, in the absence of DNA directly extracted from a skeleton with this morphology, our results are not sufficient to conclude that the Population Y we have reconstructed from the genetic data had this morphology.

An open question is when and how Population Y ancestry reached South America. There are several archaeological sites in the Americas that are contemporary to or earlier than Clovis sites. The fact that the one individual from a Clovis context who has yielded ancient DNA had entirely First American ancestry³ suggests the possibility that Population Y ancestry may be found in non-Clovis sites. Regardless of the archaeological associations, our results suggest that the genetic ancestry of Native Americans from Central and South America cannot be due to a single pulse of migration south of the Late Pleistocene ice sheets from a homogenous source population, and instead must reflect at least two streams of migration or alternatively a long drawn out period of gene flow from a structured Beringian or Northeast Asian source. The arrival of Population Y ancestry in the Americas must in any scenario have been ancient: while Population Y shows a distant genetic affinity to Andamanese, Australian and New Guinean populations, it is not particularly closely related to any of them, suggesting that the source of population Y in Eurasia no longer exists; furthermore, we detect no long-range admixture linkage disequilibrium in Amazonians as would be expected if the Population Y migration had occurred within the last few thousand years. Further insight into the population movements responsible for these findings should be

(the remainder of which is related to Mesoamerican Mixe). **b**, Two-dimensional grid of combinations of the admixture proportions α and γ which are compatible with the data in terms of how many predicted f_4 -statistics deviate by $Z \geq 3.0$ from empirical values.

possible through genome-wide analysis of ancient remains from across the Americas.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 5 February; accepted 14 July 2015.

Published online 21 July 2015.

- Wang, S. *et al.* Genetic variation and population structure in Native Americans. *PLoS Genet.* **3**, e185 (2007).
- Reich, D. *et al.* Reconstructing Native American population history. *Nature* **488**, 370–374 (2012).
- Rasmussen, M. *et al.* The genome of a Late Pleistocene human from a Clovis burial site in western Montana. *Nature* **506**, 225–229 (2014).
- Raghavan, M. *et al.* Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. *Nature* **505**, 87–91 (2014).
- Neves, W. & Pucciarelli, H. The origins of the first Americans—an analysis based on the cranial morphology of early South American remains. *Am. J. Phys. Anthropol.* **81**, 274 (1990).
- Neves, W. *et al.* Early Holocene human skeletal remains from Cerca Grande, Lagoa Santa, Central Brazil, and the origins of the first Americans. *World Archaeol.* **36**, 479–501 (2004).
- Neves, W. A., Prous, A., González-José, R., Kipnis, R. & Powell, J. Early Holocene human skeletal remains from Santana do Riacho, Brazil: implications for the settlement of the New World. *J. Hum. Evol.* **45**, 19–42 (2003).
- González-José, R. *et al.* Late Pleistocene/Holocene craniofacial morphology in Mesoamerican Paleoindians: implications for the peopling of the New World. *Am. J. Phys. Anthropol.* **128**, 772–780 (2005).
- Rasmussen, M. *et al.* Ancient human genome sequence of an extinct Palaeo-Eskimo. *Nature* **463**, 757–762 (2010).
- Raghavan, M. *et al.* The genetic prehistory of the New World Arctic. *Science* **345**, <http://dx.doi.org/10.1126/science.1255832> (2014).
- Gilbert, M. T. P. *et al.* DNA from pre-Clovis human coprolites in Oregon, North America. *Science* **320**, 786–789 (2008).
- Chatters, J. C. *et al.* Late Pleistocene human skeleton and mtDNA link Paleoamericans and modern Native Americans. *Science* **344**, 750–754 (2014).
- Jantz, R. L. & Owsley, D. W. Variation among early North American crania. *Am. J. Phys. Anthropol.* **114**, 146–155 (2001).
- Neves, W. A., Hubbe, M. & Correal, G. Human skeletal remains from Sabana de Bogota, Colombia: a case of Paleoamerican morphology late survival in South America? *Am. J. Phys. Anthropol.* **133**, 1080–1098 (2007).
- González-José, R. *et al.* Craniometric evidence for Palaeoamerican survival in Baja California. *Nature* **425**, 62–65 (2003).
- Sparks, C. S. & Jantz, R. L. A reassessment of human cranial plasticity: Boas revisited. *Proc. Natl Acad. Sci. USA* **99**, 14636–14639 (2002).
- Relethford, J. H. Apportionment of global human genetic diversity based on craniometrics and skin color. *Am. J. Phys. Anthropol.* **118**, 393–398 (2002).
- Patterson, N. *et al.* Ancient admixture in human history. *Genetics* **192**, 1065–1093 (2012).
- Lazaridis, I. *et al.* Ancient human genomes suggest three ancestral populations for present-day Europeans. *Nature* **513**, 409–413 (2014).
- Qin, P. & Stoneking, M. Denisovan ancestry in East Eurasian and Native American populations. *Mol. Biol. Evol.* (2015).
- Green, R. E. *et al.* A draft sequence of the Neandertal genome. *Science* **328**, 710–722 (2010).
- Meyer, M. *et al.* A high-coverage genome sequence from an Archaic Denisovan individual. *Science* **338**, 222–226 (2012).

23. Prüfer, K. *et al.* The complete genome sequence of a Neanderthal from the Altai Mountains. *Nature* **505**, 43–49 (2014).
24. McVicker, G., Gordon, D., Davis, C. & Green, P. Widespread genomic signatures of natural selection in hominid evolution. *PLoS Genet.* **5**, e1000471 (2009).
25. Gillespie, J. H. Genetic drift in an infinite population: the pseudohitchhiking model. *Genetics* **155**, 909–919 (2000).
26. Coop, G. *et al.* The role of geography in human adaptation. *PLoS Genet.* **5**, e1000500 (2009).
27. Moorjani, P. *et al.* The history of African gene flow into Southern Europeans, Levantines, and Jews. *PLoS Genet.* **7**, e1001373 (2011).
28. Hellenthal, G. *et al.* A genetic atlas of human admixture history. *Science* **343**, 747–751 (2014).
29. Sankararaman, S., Patterson, N., Li, H., Pääbo, S. & Reich, D. The date of interbreeding between Neandertals and modern humans. *PLoS Genet.* **8**, e1002947 (2012).
30. Lawson, D. J., Hellenthal, G., Myers, S. & Falush, D. Inference of population structure using dense haplotype data. *PLoS Genet.* **8**, e1002453 (2012).

Supplementary Information is available in the online version of the paper.

Acknowledgements We are grateful to the Native American volunteers who contributed the DNA samples used to generate the new data reported in this study and to the Fundação Nacional do Índio (FUNAI, Brazil) for logistical support in sample collection. We thank W. Klitz and C. Winkler for sharing samples for whole-genome sequencing. We thank L. Fehren-Schmitz, Q. Fu, G. Hellenthal, A. Kim, I. Lazaridis,

M. Lipson, I. Mathieson, D. Meltzer, P. Moorjani and J. Pickrell for comments, and A. Tandon for technical assistance. We thank T. Ferraz and R. Bisso-Machado for assistance with DNA extraction for the genotyping of Brazilian samples. We performed whole-genome sequencing as part of the Simons Genome Diversity Project. Genotyping of the Brazilian samples was performed at the Children's Hospital of Philadelphia and we particularly thank C. Hou for her support in this. M.C.B., T.H., M.L.P.-E. and F.M.S. were supported by Conselho Nacional do Desenvolvimento Científico e Tecnológico and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (Brazil). P.S. was supported by the Wenner-Gren foundation and the Swedish Research Council (VR grant 2014-453). D.R. was supported by US National Science Foundation HOMINID grant BCS-1032255, US National Institutes of Health grant GM100233, Simons Foundation Grant 280376 and the Howard Hughes Medical Institute.

Author Contributions P.S. performed analyses. P.S., S.M., M.C.B., N.C., T.H., M.L.P.-E., F.M.S., N.P. and D.R. prepared datasets. P.S. and D.R. wrote the paper.

Author Information Genome sequence data is available from (<https://www.simonsfoundation.org/life-sciences/simons-genome-diversity-project-dataset/>). New Affymetrix Human Origins array genotype data are available to researchers who send D.R. a signed letter agreeing to respect specific conditions (Supplementary Information section 1). Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to P.S. (skoglund@genetics.med.harvard.edu) or D.R. (reich@genetics.med.harvard.edu)

METHODS

Data reporting. No statistical methods were used to predetermine sample size. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment.

New Affymetrix Human Origins genotypes. We generated new Affymetrix Human Origins array genotypes for 48 individuals from 9 populations from present-day Brazil (Apalaí, Arara, Guarani_GN, Guarani_KW, Karitiana, Suruí, Urubu Kaapor, Xavante and Zoró). Ethical approval for the sample collection was provided by the Brazilian National Ethics Commission (CONEP Resolution no. 123/98). CONEP also approved the oral consent procedure and the use of these samples in studies of population history and human evolution. Individual and/or tribal informed oral consents were obtained from participants who were not able to read or write. All sampling was coordinated by co-authors of this study (M.L.P.-E. and F.M.S.) and their collaborators, in a manner consistent with the Helsinki Declaration and Brazilian laws and regulations applicable at the time of sampling. Logistical support for the sample collection was provided by the Fundação Nacional do Índio (FUNAI). We curated the data in the same way as reported in ref. 19 (Supplementary Information section 1). We computationally phased these data together with the previously published Affymetrix Human Origins SNP array data using SHAPEIT2 (ref. 31) with default parameters.

High-coverage genome sequencing and processing. We sent samples from 18 Papuan, Mixe, Suruí and Yoruba individuals to Illumina for deep-coverage sequencing using a non-PCR-based protocol as part of the Simons Genome Diversity Project. The sequence reads were mapped using the 'aln' algorithm of BWA (version 0.5.10)³² and genotypes were inferred using the unified genotyper from GATK³³ (version 2.5.2-gf57256b). These data are available from (<https://www.simonfoundation.org/life-sciences/simons-genome-diversity-project-dataset/>). Briefly, sequence reads were stripped of adapters before alignment to the decoy version of the hg19 reference sequence (hs37d5). Read groups were added for identification and compatibility with GATK tools, before indel realignment and duplicate removal. The genotyping performed thereafter used a reference-free procedure that reduces reference bias. A specially developed filtering engine assigned filtering levels from 0 to 9 for each position in the genome. All population genetic analyses in this paper used the most stringent level of filtering (level 9).

Testing for more than one ancestral population of Central and South Americans. To investigate whether Central and South American populations are consistent with being derived from a single stream of ancestry, we applied the software qpWave² to ask the question whether the set of f_4 -statistics of the form $f_4(A = \text{American}_1, B = \text{American}_2; X = \text{outgroup}_1, Y = \text{outgroup}_1) = (p_A - p_B)(p_X - p_Y)$ forms a matrix that is consistent with being of rank 0 (averaged over all SNPs, where p_A, p_B, p_X , and p_Y are the frequencies of an arbitrarily chosen allele in populations A, B, X and Y at each locus). If all these Native American populations descend from the same stream of migration into the Americas, then the f_4 -statistic relating each Native American population to each non-Native American population should be the same for all Native American populations, and in particular consistent with 0. Formally, to evaluate whether the f_4 -statistic matrix is consistent with being of rank 0, we compute a Hotelling's T^2 test that appropriately corrects for the correlation structure of the f_4 -statistics. We analysed 7 Native American populations each with at least 3 individuals with no detected post-Columbian admixture, and 4 populations from each of 6 worldwide regions as outgroups (Supplementary Information section 2).

D-statistic tests based on correlation in allele frequencies. To investigate whether a tree-like population history ((A, B), (X, Y)) is consistent with the data, for example, with A = chimpanzee, B = Onge, X = Mixe and Y = Suruí, we computed D-statistics^{18,21}

$$D(A, B; X, Y) = \frac{(p_A - p_B)(p_X - p_Y)}{(p_A + p_B - 2p_{AB})(p_X + p_Y - 2p_{XY})}$$

over all SNPs, where p_A, p_B, p_X , and p_Y are the frequencies of an arbitrarily chosen allele in populations A, B, X and Y at each locus. We computed standard errors using a block jackknife weighted by the number of SNPs in each 5 cM (5 Mb in the case of high-coverage genome sequences) block in the genome^{34,35}. We report Z-scores as normalized $Z = D/\text{s.e.}$ and we interpret statistics $|Z| > 3$ as being significantly different from 0. We only considered SNPs that were informative, in the sense that they are polymorphic both within (A, B) and (X, Y).

Correlation of signal to regions of functional importance. We divided the genome into 10 deciles of the 'B-value' described in ref. 24, which integrates multiple genomic annotations into a single estimate of proximity to functional regions for each nucleotide in the genome. We then used linear regression to estimate the coefficient a of the function $y = ax + c$ where $x = B$ (the rank of the decile of B) and $y = D_B$ (D restricted to the particular decile of B). To compute standard errors, we used a weighted block jackknife procedure where

each 5 Mb block of the genome is dropped in turn and a is recomputed. The variability of a across each of these leave-one-out computations, weighting by the number of informative loci in each block, was what we used to estimate a standard error^{34,35}.

h_4 -statistic tests based on correlation in linkage disequilibrium. We devised a linkage disequilibrium statistic that tests for symmetry in linkage disequilibrium between two proposed clades with a pair of populations in each. The statistic, h_4 , is:

$$h_4 = ((p_{12}^A - p_1^A p_2^A) - (p_{12}^B - p_1^B p_2^B)) \times ((p_{12}^C - p_1^C p_2^C) - (p_{12}^D - p_1^D p_2^D))$$

where 1 and 2 are arbitrarily chosen reference alleles at two different loci, respectively, and A, B, C, and D denote four different populations. Thus, p_{12}^A is the frequency of the 12 haplotype in population A, and p_1^A is the frequency of the 1 allele in population A. The quantity $p_{12}^A - p_1^A p_2^A$ thus measures the difference between the observed haplotype frequency and the expected haplotype frequency given the allele frequencies³⁶. The motivation for this statistic being informative about population history is that under a tree-like model ((A, B), (C, D)) with no gene flow, differences in linkage disequilibrium between populations A and B are not expected to correlate to differences in linkage disequilibrium between populations C and D. If there has been gene flow between the two clades, the statistic may be significantly positive or negative like f_4 - and D-statistics¹⁸.

In practice, we computed this statistic for each polymorphic locus ('target locus') by identifying all other polymorphic loci 5' of the target locus at distance interval $d \pm w$ and computing the statistic for each pairing. We then averaged the statistic over all valid pairs of loci in the genome identified in this way. We computed standard errors using a block jackknife over contiguous 5 cM blocks in the genome, where SNP pairs that bridge the boundary of two blocks are assigned to the block in which the target locus is found. For the main analysis we computed h_4 -statistics of the form $h_4(\text{Yoruba}, X; \text{Mixe}, \text{Suruí})$ for all populations X genotyped using the Affymetrix Human Origins SNP array, and all pairs of SNPs within 0.01 cM of each other. We restricted the analysis to populations with at least 10 individuals. We also computed the h_4 -statistic for windows of 0.001 cM centred around different genetic distances for selected populations (Extended Data Fig. 5).

Chromosome painting symmetry tests. We used SHAPEIT to phase 593,142 SNPs with the same set of individuals as described above, using all autosomal SNPs in the Affymetrix Human Origins array. We then 'painted' unadmixed Native American individuals using non-American populations, and excluded the Yukagir and the Chukchi since they have evidence of back-migration from the Americas. We ran ChromoPainter v2 using default parameters, painting each recipient individual separately, but using all donor populations as candidates to paint each recipient haplotype. To assess statistical uncertainty, we repeated this procedure for each recipient individual using 22 subsets of the data where for each of these subsets a different autosome had been dropped. We then used the results of these 22 block jackknife pseudo-replicates to obtain a weighted block jackknife estimate of the standard error for our test statistic (see below).

To test if the recipient populations copied equally from the donor populations, we computed the average 'chunk count' $C_{R,D}$ copied from a given donor population D in each recipient population R (averaged over individuals). We then computed a $S(R_1, R_2; D)$ statistic that quantifies the symmetry between two Native American populations in their copying from each donor:

$$S(D; R_2, R_1) = \frac{C_{R_1,D} - C_{R_2,D}}{C_{R_1,D} + C_{R_2,D}}$$

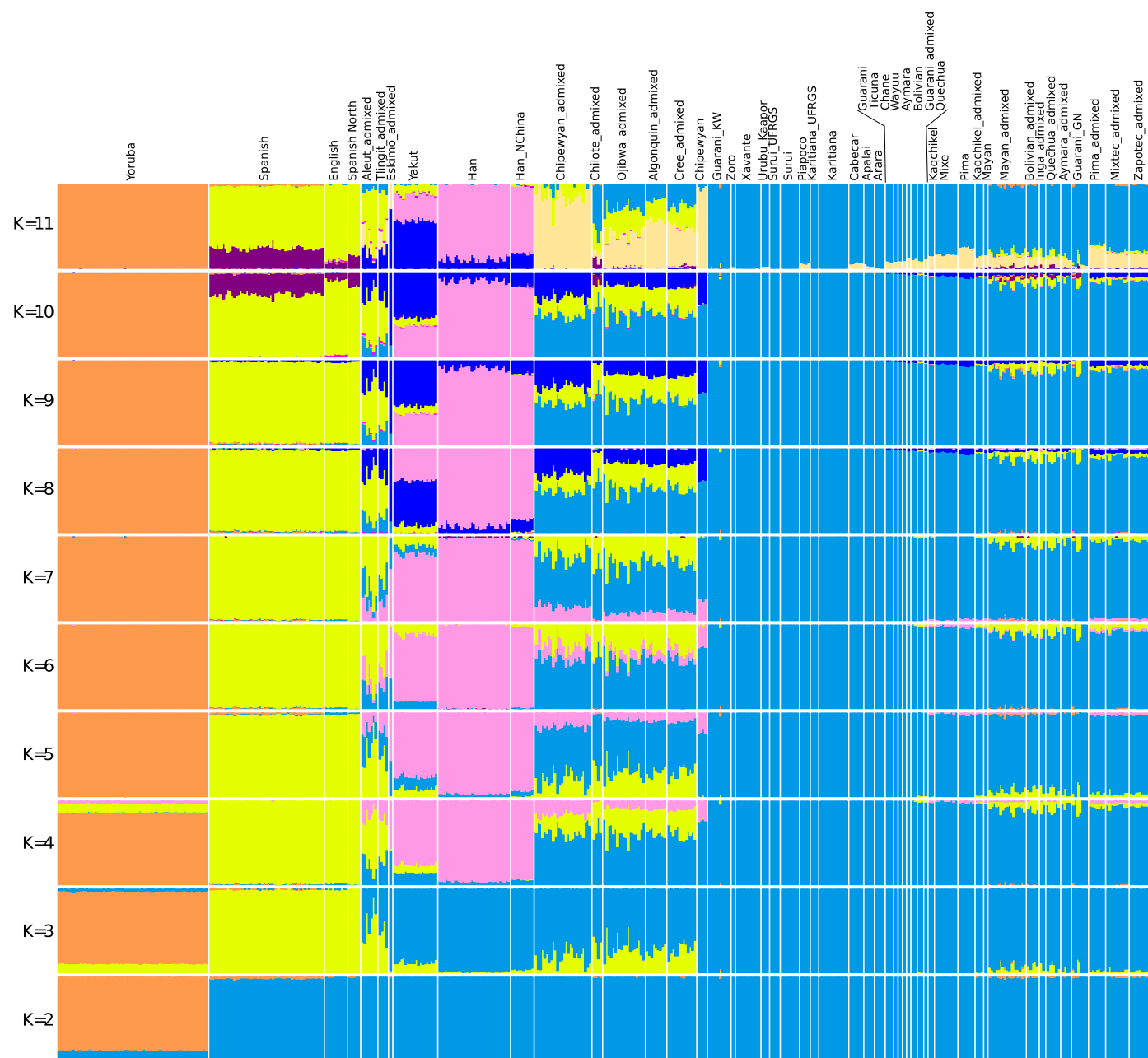
If two Native American populations, such as the Suruí and the Mixe, derive all of their ancestry from a single common origin, we expect that they would copy from the donor populations at an equal rate. We computed the standard error of this statistic using the 22 subsets of the data where each autosome had been dropped, weighted using the number of SNPs on each chromosome. We generated the world map in Fig. 1d by using the R maps package to plot the value of $S(X; \text{Mixe}, \text{Suruí}; \text{Karitiana})$ for each non-American population X, and $S(\text{Onge}; \text{Mixe}, Y)$ for each American population Y.

Admixture graph models of population relationships. We used ADMIXTURE-GRAPH¹⁸ to fit suggested phylogenies with admixture events to the data. We assessed goodness-of-fit by investigating all possible f -statistics predicted by the fitted model and assessing whether they differed significantly from the empirical data. We chose as a starting point the model relating Mbuti Africans, Andamanese Onge, MA1 and Karitiana fitted by a previous study¹⁹ where lineages related to MA1 and the Onge both contributed ancestry to the Karitiana. We added to this Han Chinese to represent a population that is phylogenetically more closely related to one of the ancestral populations of Native Americans than are the Onge (Extended Data Figs 6 and 7). We find that this model is inconsistent with the data, as the model predicts that Mixe and Suruí/Karitiana are equally related to Onge, and indeed we observe several statistics for which the Z-score for the

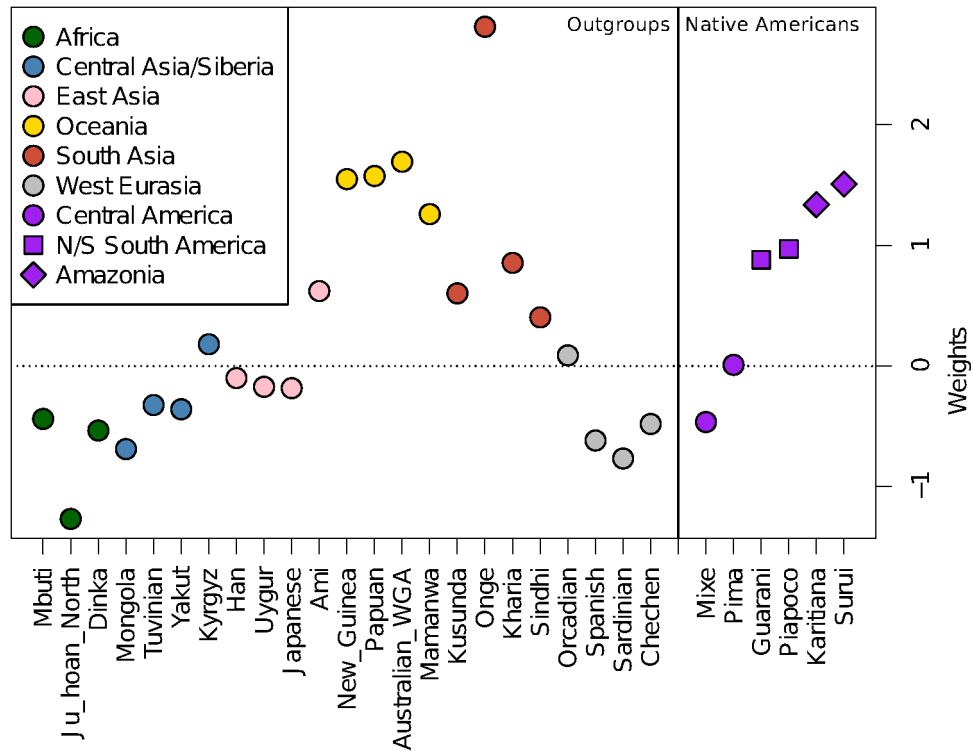
difference between the predicted and empirical statistics is $|Z| > 3$ (Extended Data Table 3). To account for this, we fitted a model in which the ancestors of Amazonians received admixture from a population related to the Onge (Extended Data Fig. 6), and found that this provides an excellent fit to the data, with no $|Z|$ -score differences greater than 3. In contrast, alternative models of Han-related or MA1-related gene flow into the Americas are inconsistent with the data (Extended Data Fig. 6 and Extended Data Table 3).

Code availability. A python program for computing h_4 symmetry statistics and other population genetic statistics used in this paper is available at (<https://github.com/pontuskk/popstats>).

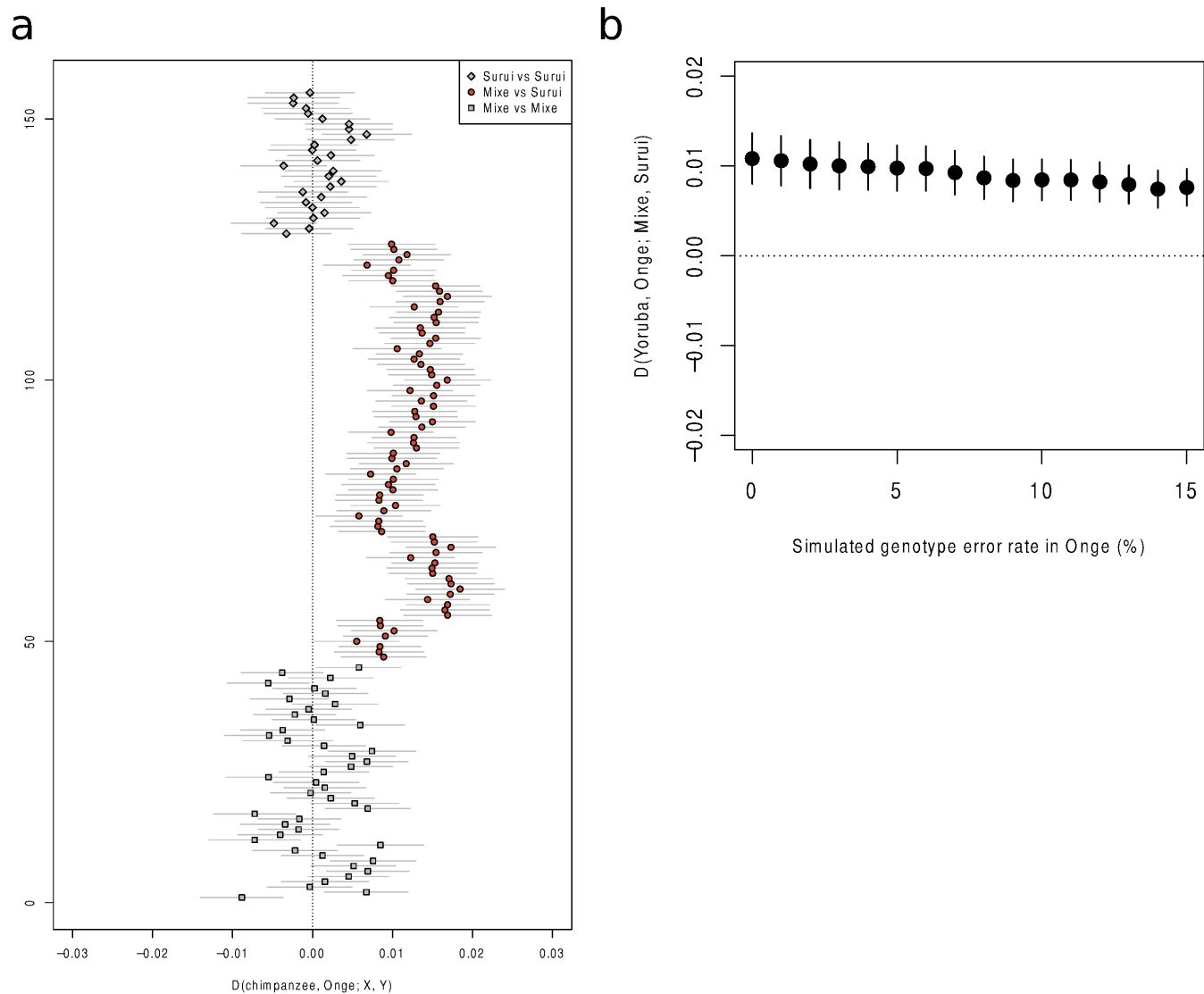
31. Delaneau, O., Marchini, J. & Zagury, J.-F. A linear complexity phasing method for thousands of genomes. *Nature Methods* **9**, 179–181 (2011).
32. Li, H. & Durbin, R. Fast and accurate short read alignment with Burrows–Wheeler transform. *Bioinformatics* **25**, 1754–1760 (2009).
33. McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* **20**, 1297–1303 (2010).
34. Busing, F. M., Meijer, E. & Van Der Leeden, R. Delete-m jackknife for unequal m. *Stat. Comput.* **9**, 3–8 (1999).
35. Reich, D., Thangaraj, K., Patterson, N., Price, A. L. & Singh, L. Reconstructing Indian population history. *Nature* **461**, 489–494 (2009).
36. Robbins, R. B. Some applications of mathematics to breeding problems III. *Genetics* **3**, 375–389 (1918).
37. Becker, R. A. & Wilks, A. R. Maps in S. *AT&T Bell Laboratories Statistics Research Report [93.2]*, (1993).
38. Alexander, D. H., Novembre, J. & Lange, K. Fast model-based estimation of ancestry in unrelated individuals. *Genome Res.* **19**, 1655–1664 (2009).



Extended Data Figure 1 | Clustering analysis. ADMIXTURE³⁸ clustering analysis performed on the Affymetrix Human Origins data used in this study. To aid in visualization, we only show results for Native American samples and for selected samples from Eurasian populations.



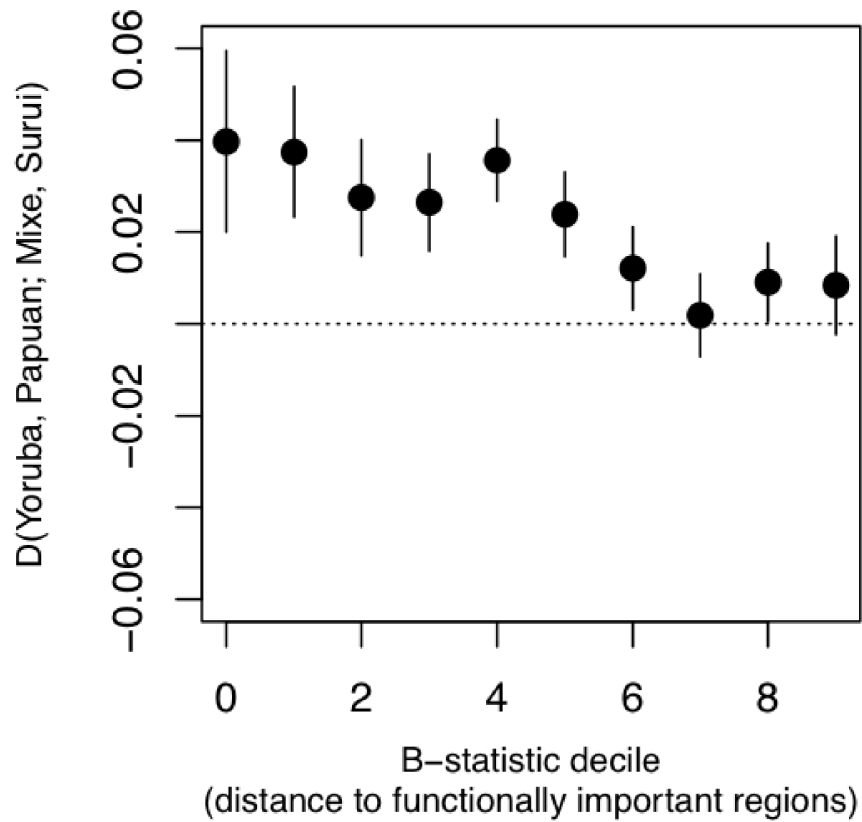
Extended Data Figure 2 | qpWave coefficients. Weights from qpWave for Native American populations and for non-American outgroup populations. No weights are given for Yoruba and Cabecar, as they are used in the computation.



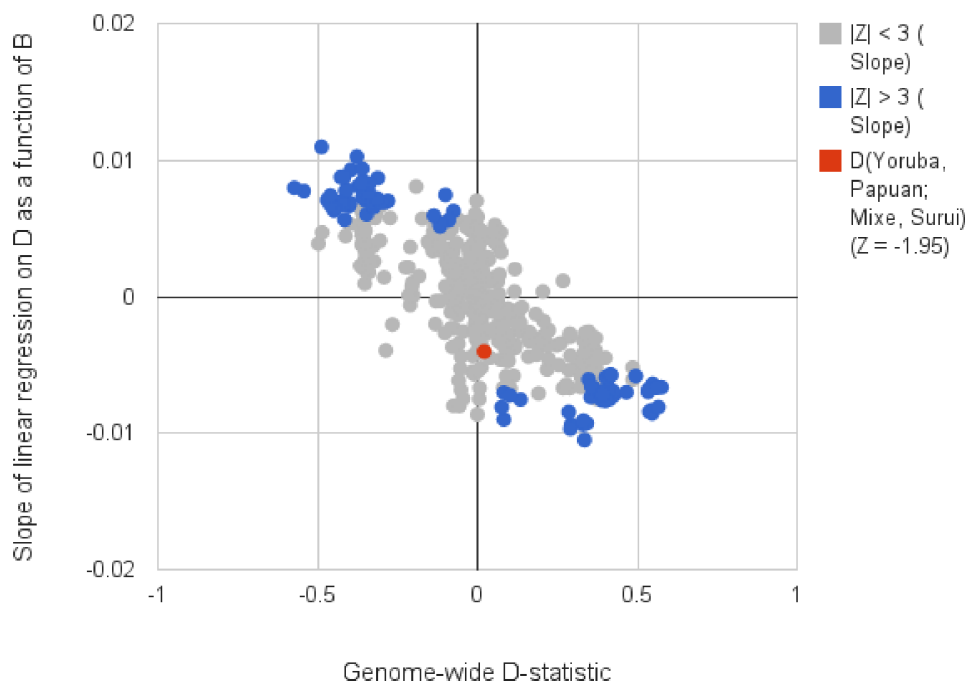
Extended Data Figure 3 | Excess allele sharing between the Surui and the Onge. **a**, Tests for excess shared derived alleles with the Onge in all possible comparisons of 8 Surui and 10 Mixe individuals. All Mixe–Surui comparisons show a positive skew whereas all Mixe–Mixe and Surui–Surui comparisons

are consistent with 0. Lines correspond to one standard error in either direction. **b**, Random sequence or genotype errors cannot explain the affinity of the Amazonians to Australasians, as simulated increased errors in the Onge do not cause an increased affinity to Surui.

a

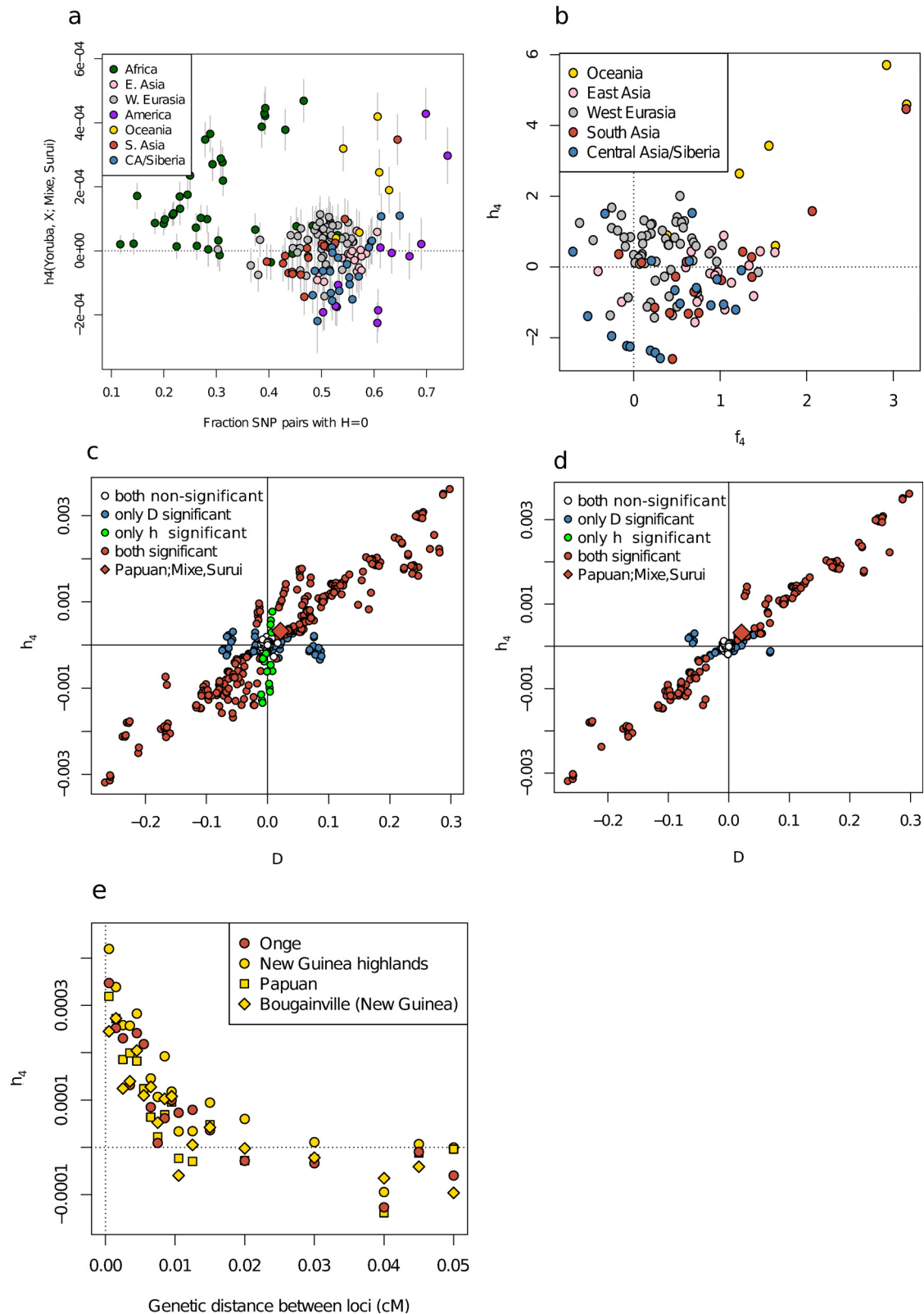


b



Extended Data Figure 4 | Signals of admixture as a function of proximity to functional regions. **a**, The affinity of 16 Papuan high-coverage genomes to 2 Amazonian Surui high-coverage genomes as a function of proximity to regions of functional importance (measured by *B*-value). **b**, A total of 395 tests

of quartets $D(\text{Yoruba}, X; Y, Z)$ shows that quartets with significantly positive slopes ($|Z| > 3$) also yield significant genome-wide *D*-statistics of the opposite sign. This suggests that signals of admixture are systematically stronger close to functionally important regions.

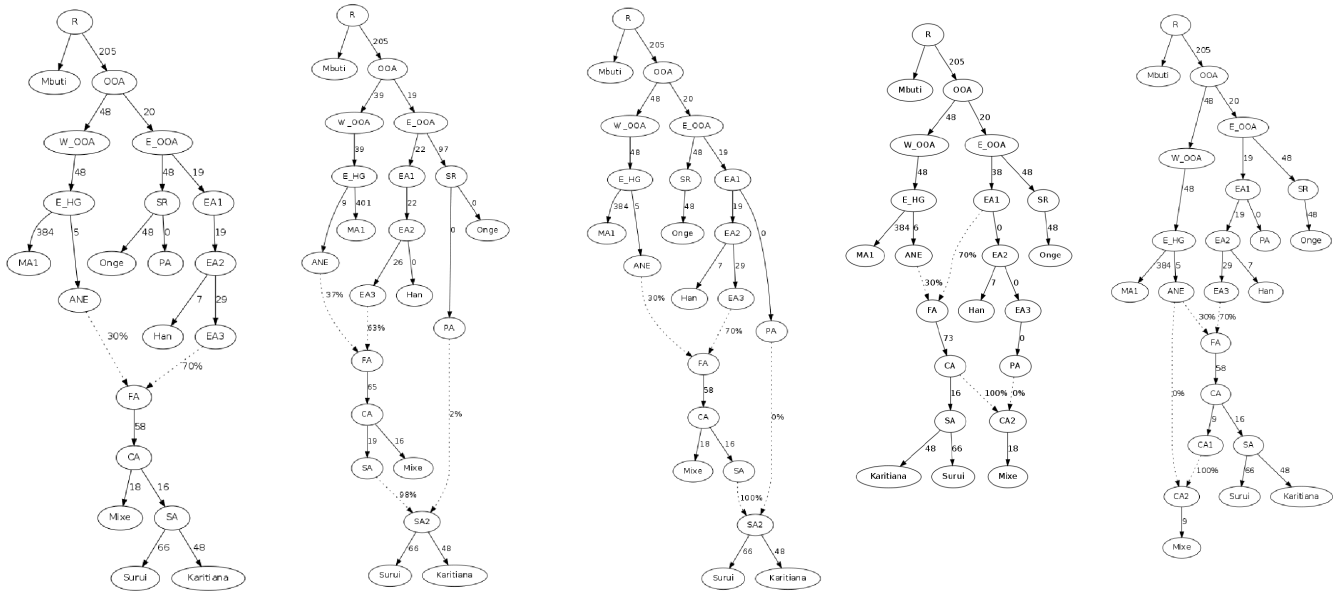


Extended Data Figure 5 | Linkage disequilibrium-based symmetry tests.

a, $h_4(\text{Yoruba, X; Mixe, Surui})$ for SNP pairs within 0.01 cM of each other contrasted with the fraction of SNP pairs in linkage equilibrium in population X ($H = 0$). Error bars show ± 1 s.e. **b**, Scatterplot of Z-scores for the f_4 - and h_4 -statistics for the same quartets. For both these panels we only use populations with at least 6 samples. **c**, **d**, We computed $D(\text{Yoruba, X; Y, Z})$ and $h_4(\text{Yoruba, X; Y, Z})$ for many combinations of populations as X, Y and Z using

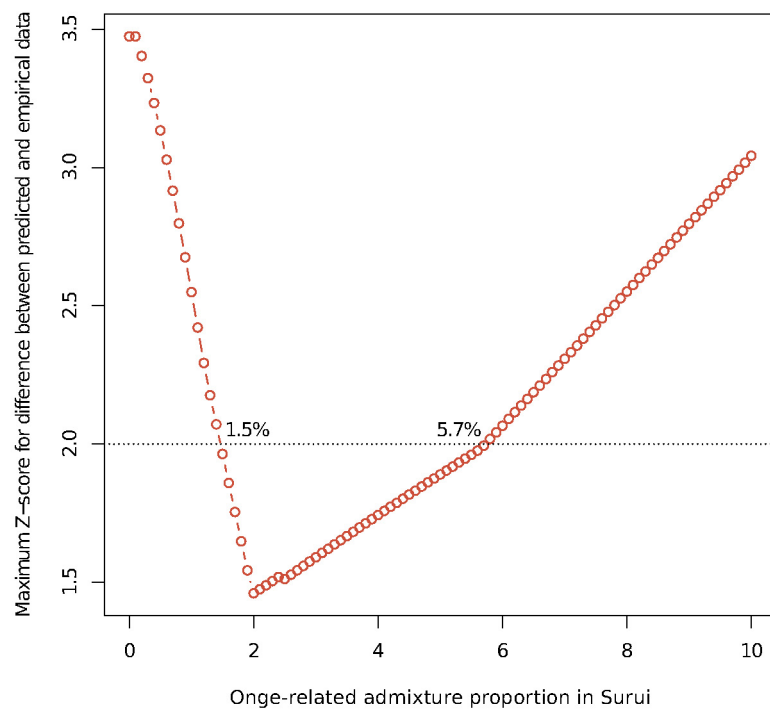
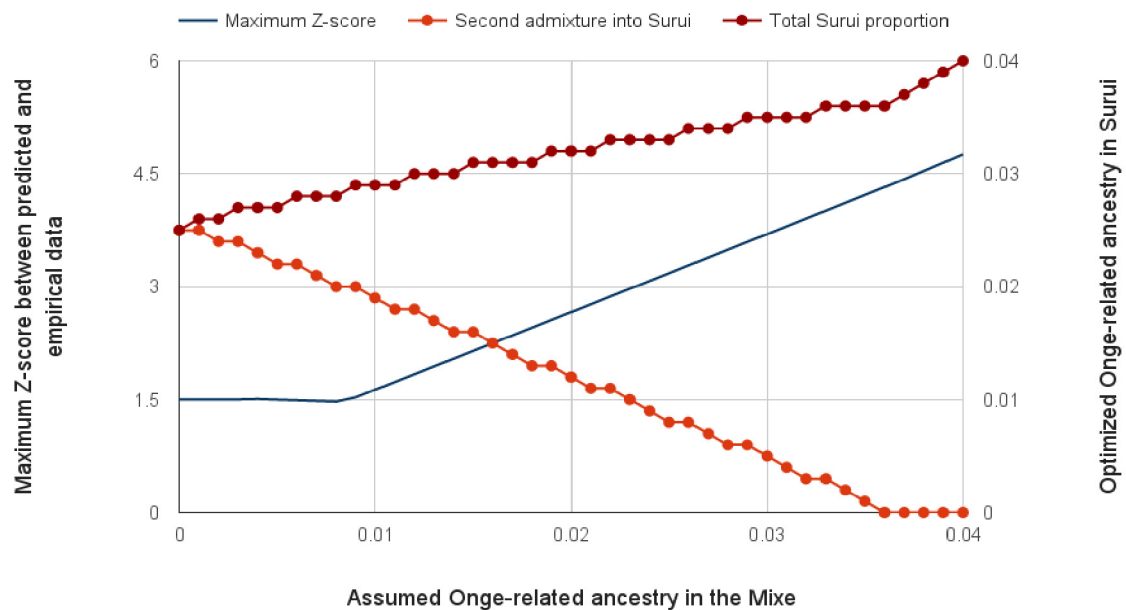
phased Affymetrix Human Origins SNP array data ascertained in a Yoruba individual. Except for Africans who have ancestry from lineages that diverged before the Yoruba used for ascertainment and Oceanians (who have archaic Denisovan ancestry), we observe that $|Z| > 3$ h_4 -statistics are always associated with a significantly positive D for the same quartet. **e**, Correlation of the h_4 -statistic with the genetic distance separation of pairs of SNPs for $h_4(\text{Yoruba, X; Mixe, Surui})$.

- a. Single origin (no fit) b. Onge-related mixture in Amazonia (fit) c. Asian mixture in Amazonia (no fit) d. Asian admixture in Mesoamerica (no fit) e. MA-1 related mixture in Mesoamerica (no fit)



Extended Data Figure 6 | Admixture graphs for fitted population history models. **a**, An admixture graph where all Mixe, Suruí and Karitiana are of 100% First American ancestry is rejected with 6 predicted f -statistics at least 3 standard errors from the empirically observed value. **b**, An admixture graph where the ancestors of Suruí and Karitiana receive 2% ancestry from a lineage related to the Onge is consistent with the data with no outliers. **c**, An admixture

graph where the distinct ancestry in Amazonians is more closely related to Han than to Onge produces 6 outliers. **d**, An admixture graph with no distinctive ancestry in Karitiana or Suruí but East Asian gene flow into the Mixe produces 7 outliers. **e**, An admixture graph with no distinctive ancestry in Karitiana or Suruí but MA1-related gene flow into the Mixe produces 6 outliers.

a**b**

Extended Data Figure 7 | Plausible range for the non-First American admixture proportion in Amazonians. a, Range obtained assuming entirely First American ancestry in the Mixe. **b,** The maximum proportion of non-First American ancestry in the Mixe that is consistent with the data.

Extended Data Table 1 | qpWave analysis provides evidence that Central and South American genetic variation is inconsistent with being derived from a single homogeneous population

	<i>P</i> -value for this number of streams			
	1	2	3	4
Full data	2.03E-07	0.09	0.58	0.92
<i>Outgroup region dropped</i>				
Africa	1.67E-04	0.34	0.92	0.95
C. Asia/Siberia	5.91E-07	0.11	0.6	0.89
East Asia	4.46E-09	0.04	0.57	0.92
South Asia	6.95E-05	0.1	0.4	0.82
West Eurasia	1.41E-05	0.06	0.37	0.89
Oceania	4.39E-05	0.43	0.88	0.97
<i>Native American population dropped</i>				
Cabecar	1.13E-08	0.02	0.27	0.73
Guarani	9.50E-07	0.27	0.76	0.99
Karitiana	1.41E-06	0.1	0.61	0.86
Mixe	8.32E-03	0.2	0.91	0.98
Piapoco	1.30E-04	0.55	0.93	0.98
Pima	2.19E-05	0.31	0.78	0.97
Surui	2.35E-06	0.11	0.56	0.84
<i>Africa + 1 other region</i>				
Siberia	0.16	0.56	0.8	0.87
East Asia	0.06	0.29	0.71	0.72
South Asia	0.004	0.57	0.93	0.96
West Eurasia	0.02	0.23	0.62	0.67
Oceania	0.01	0.25	0.82	0.99
<i>Siberia + 1 other region</i>				
Africa	0.16	0.56	0.8	0.87
East Asia	0.41	0.91	0.99	1
South Asia	0.03	0.82	0.91	0.9
W. Eurasia	0.2	0.59	0.77	0.83
Oceania	0.003	0.18	0.75	0.93

Extended Data Table 2 | Top 20 *D*-statistics observed for *D*(chimpanzee, Old World population; Central Americans, Amazonians)

Rank	Population	<i>D</i>	SE	<i>Z</i>	Region 1	Region 2
1	Onge	0.0101	0.0022	4.60	India	South Asia
2	Papuan	0.0084	0.0022	3.82	Papua New Guinea	Oceania
3	New_Guinea	0.0082	0.0023	3.54	Papua New Guinea	Oceania
4	Australian_WGA	0.0074	0.0024	3.12	Australia (Arnhem Land)	Oceania
5	Mamanwa	0.0068	0.0020	3.40	Philippines (Negrito)	Oceania
6	Bougainville	0.0065	0.0023	2.85	Papua New Guinea	Oceania
7	Kharia	0.0059	0.0020	2.97	India	South Asia
8	Tongan	0.0058	0.0022	2.68	Tonga	Oceania
9	Bengali	0.0058	0.0019	3.00	Bangladesh	South Asia
10	Mala	0.0055	0.0019	2.93	India	South Asia
11	Ami	0.0052	0.0020	2.61	Taiwan	East Asia
12	Lodhi	0.0052	0.0019	2.72	India	South Asia
13	Sindhi	0.0051	0.0019	2.72	Pakistan	South Asia
14	Kusunda	0.0050	0.0020	2.56	Nepal	South Asia
15	Lahu	0.0050	0.0021	2.37	China	East Asia
16	Kinh	0.0049	0.0020	2.46	Vietnam	East Asia
17	Australian	0.0048	0.0025	1.96	Australia	Oceania
18	Balochi	0.0047	0.0019	2.55	Pakistan	South Asia
19	Thai	0.0047	0.0020	2.38	Thailand	East Asia
20	Semende	0.0045	0.0020	2.27	Indonesia (Sumatra)	Oceania

Extended Data Table 3 | f_4 -statistics for which the statistic predicted by the fitted admixture graphs deviates by more than $|Z| > 3$ from the statistic computed on the empirical data

<i>A</i>	<i>B</i>	<i>X</i>	<i>Y</i>	Predicted f_4	Empirical f_4	Z-score
<i>Single First American origin (Extended Data Figure 6A)</i>						
Mbuti	Onge	Mixe	Surui	0	0.003506	3.535
Mbuti	Onge	Mixe	Karitiana	0	0.00315	3.431
Onge	Mixe	Mixe	Surui	-0.018466	-0.021724	-3.061
Onge	Mixe	Mixe	Karitiana	-0.018466	-0.021849	-3.226
Onge	Han	Mixe	Surui	0	-0.002902	-3.654
Onge	Han	Mixe	Karitiana	0	-0.00239	-3.279
<i>Onge-related ancestry in the Amazon (Extended Data Figure 6B)</i> (No outliers)						
<i>East Asian admixture in South America (Extended Data Figure 6C)</i>						
Mbuti	Onge	Mixe	Surui	0	0.003506	3.535
Mbuti	Onge	Mixe	Karitiana	0	0.00315	3.431
Onge	Mixe	Mixe	Surui	-0.018466	-0.021724	-3.061
Onge	Mixe	Mixe	Karitiana	-0.018466	-0.021849	-3.226
Onge	Han	Mixe	Surui	0	-0.002902	-3.654
Onge	Han	Mixe	Karitiana	0	-0.00239	-3.279
<i>East Asian admixture in Central America (Extended Data Figure 6D)</i>						
Mbuti	Onge	Mixe	Surui	-0.000002	0.003506	3.537
Mbuti	Onge	Mixe	Karitiana	-0.000002	0.00315	3.433
Onge	Mixe	Mixe	Surui	-0.018466	-0.021724	-3.061
Onge	Mixe	Mixe	Karitiana	-0.018466	-0.021849	-3.225
Onge	Han	Mixe	Surui	-0.000004	-0.002902	-3.649
Onge	Han	Mixe	Karitiana	-0.000004	-0.00239	-3.273
<i>Ancient Siberian (MA1) admixture in Central America (Extended Data Figure 6E)</i>						
Mbuti	Onge	Mixe	Surui	0	0.003506	3.535
Mbuti	Onge	Mixe	Karitiana	0	0.00315	3.431
Onge	Mixe	Mixe	Surui	-0.018470	-0.021724	-3.057
Onge	Mixe	Mixe	Karitiana	-0.018470	-0.021849	-3.222
Onge	Han	Mixe	Surui	0	-0.002902	-3.654
Onge	Han	Mixe	Karitiana	0	-0.00239	-3.279

Mutations in *DCHS1* cause mitral valve prolapse

Ronen Durst^{1,2*}, Kimberly Sauls^{3*}, David S. Peal^{4*}, Annemarieke deVlaming³, Katelynn Toomer³, Maire Leyne¹, Monica Salani¹, Michael E. Talkowski^{1,5}, Harrison Brand^{1,5}, Maëlle Perrocheau⁶, Charles Simpson¹, Christopher Jett¹, Matthew R. Stone¹, Florie Charles¹, Colby Chiang¹, Stacey N. Lynch⁴, Nabila Bouatia-Naji^{6,7}, Francesca N. Delling⁸, Lisa A. Freed⁹, Christophe Tribouilloy¹⁰, Thierry Le Tourneau¹¹, Hervé LeMarec¹¹, Leticia Fernandez-Friera^{12,13}, Jorge Solis^{12,13}, Daniel Trujillano^{14,15,16,17}, Stephan Ossowski^{15,18}, Xavier Estivill^{14,15,16,17}, Christian Dina^{11,19,20,21}, Patrick Bruneval²², Adrian Chester²³, Jean-Jacques Schott^{11,19,20,21}, Kenneth D. Irvine²⁴, Yaopan Mao²⁴, Andy Wessels³, Tahirali Motiwala³, Michel Puceat²⁵, Yoshikazu Tsukasaki²⁶, Donald R. Menick²⁷, Harinath Kasiganesan²⁷, Xingju Nie²⁸, Ann-Marie Broome²⁸, Katherine Williams³, Amanda Johnson³, Roger R. Markwald³, Xavier Jeunemaitre^{6,7,29}§, Albert Hagege^{6,7,30}§, Robert A. Levine^{7,31}§, David J. Milan^{1,4}§, Russell A. Norris³§ & Susan A. Slaugenhaupt¹§

Mitral valve prolapse (MVP) is a common cardiac valve disease that affects nearly 1 in 40 individuals^{1–3}. It can manifest as mitral regurgitation and is the leading indication for mitral valve surgery^{4,5}. Despite a clear heritable component, the genetic aetiology leading to non-syndromic MVP has remained elusive. Four affected individuals from a large multigenerational family segregating non-syndromic MVP underwent capture sequencing of the linked interval on chromosome 11. We report a missense mutation in the *DCHS1* gene, the human homologue of the *Drosophila* cell polarity gene *dachsous* (*ds*), that segregates with MVP in the family. Morpholino knockdown of the zebrafish homologue *dachsous1b* resulted in a cardiac atrioventricular canal defect that could be rescued by wild-type human *DCHS1*, but not by *DCHS1* messenger RNA with the familial mutation. Further genetic studies identified two additional families in which a second deleterious *DCHS1* mutation segregates with MVP. Both *DCHS1* mutations reduce protein stability as demonstrated in zebrafish, cultured cells and, notably, in mitral valve interstitial cells (MVICs) obtained during mitral valve repair surgery of a proband. *Dchs1*^{+/-} mice had prolapse of thickened mitral leaflets, which could be traced back to developmental errors in valve morphogenesis. *DCHS1* deficiency in MVP patient MVICs, as well as in *Dchs1*^{+/-} mouse MVICs, result in altered migration and cellular patterning, supporting these processes as aetiological underpinnings for the disease. Understanding the role of *DCHS1* in mitral valve development and MVP pathogenesis holds potential for therapeutic insights for this very common disease.

In a previous study, based on specific diagnostic criteria^{6–9}, *MMVP2* (myxomatous mitral valve prolapse-2) was mapped to a 4.3 cM region of chromosome 11p15.4 in a family of Western European descent

segregating non-syndromic mitral valve prolapse as an autosomal dominant trait with age-dependent penetrance (Fig. 1a, c)⁶. We performed tiled capture and high-throughput sequence analysis of genomic DNA from four affected individuals (Fig. 1a), identifying 4,891 single nucleotide variants (SNVs) and insertion/deletion polymorphisms in the targeted region (see Methods). After selecting rare protein-coding variants shared among all affected pedigree members, we identified three heterozygous protein-altering variants: two missense SNVs in *DCHS1*, a member of the cadherin superfamily¹⁰, resulting in p.P197L and p.R2513H (Fig. 1b), and a single missense SNV in *APBB1*, the amyloid beta (A4) precursor protein-binding family B, member 1 gene resulting in p.R481H. Both *DCHS1* mutations, p.P197L and p.R2513H, were rare in the population (the former observed three times in 4,300 European-American individuals from the NHLBI Exome Sequencing Project and the latter never observed), and both were predicted to be protein damaging by PolyPhen-2 (ref. 11), LRT¹², and MutationTaster¹³. While the *APBB1* variant was also rare in population-based data, no cardiac phenotype was observed in *apbb1* morphant zebrafish, despite reduction of *apbb1* mRNA (Extended Data Fig. 1a, b). Additionally, *Apbb1* is not expressed in murine cardiac valves (Extended Data Fig. 2)¹⁴, and no cardiac defects have been reported in the *Apbb1* knockout mouse¹⁵. This suggests that the *APBB1* variant is unlikely to be contributing to MVP in this family.

The functional effect of the *DCHS1* variants was evaluated in the zebrafish *Danio rerio*, as this model system lends itself to functional annotation of mutations implicated in human disease^{16–18}. Zebrafish have two *DCHS1* homologues, *dachsous1a* and *dachsous1b*. *dachsous1b* is located in a region of *D. rerio* chromosome 10 that is syntenic to the *DCHS1* region of human chromosome 11. Knockdown of *dachsous1a* did not result in a cardiac phenotype despite reduction in mRNA levels

¹Center for Human Genetic Research, Massachusetts General Hospital Research Institute and Department of Neurology, Harvard Medical School, 185 Cambridge Street, Boston, Massachusetts 02114 USA. ²Cardiology Division, Hadassah Hebrew University Medical Center, POB 12000 Jerusalem, Israel. ³Cardiovascular Developmental Biology Center, Department of Regenerative Medicine and Cell Biology, Department of Medicine, Children's Research Institute, Medical University of South Carolina, 171 Ashley Avenue, Charleston, South Carolina 29425, USA. ⁴Cardiovascular Research Center, Cardiology Division, Massachusetts General Hospital, Harvard Medical School, 55 Fruit Street, Boston, Massachusetts 02114, USA. ⁵Psychiatric and Neurodevelopmental Genetics Unit, Department of Psychiatry, Massachusetts General Hospital, Boston, Massachusetts 02114, USA. ⁶INSERM, UMR-970, Paris Cardiovascular Research Center, 75015 Paris, France. ⁷Université Paris Descartes, Sorbonne Paris Cité, Faculty of Medicine, 75006 Paris, France. ⁸Department of Medicine (Cardiovascular Division), Beth Israel Deaconess Medical Center, Harvard Medical School, Boston, Massachusetts 02215, USA. ⁹Yale-New Haven Hospital Heart and Vascular Center, Yale School of Medicine, 20 York Street, New Haven, Connecticut 06510, USA. ¹⁰Department of Cardiology, University Hospital Amiens; INSERM U-1088, Jules Verne University of Picardie, 80000 Amiens, France. ¹¹Inserm U1087; Institut du Thorax; University Hospital, 44007 Nantes, France. ¹²Centro Nacional de Investigaciones Cardiovasculares, Carlos III (CNIC), 28029 Madrid, Spain. ¹³Hospital Universitario Montepíncipe, 28660 Madrid, Spain. ¹⁴Genetic Causes of Disease Group, Centre for Genomic Regulation (CRG), 08003 Barcelona, Catalonia, Spain. ¹⁵Universitat Pompeu Fabra (UPF), 08002 Barcelona, Catalonia, Spain. ¹⁶Hospital del Mar Medical Research Institute (IMIM), 08003 Barcelona, Catalonia, Spain. ¹⁷CIBER in Epidemiology and Public Health (CIBERESP), 08036 Barcelona, Catalonia, Spain. ¹⁸Genomic and Epigenomic Variation in Disease Group, Centre for Genomic Regulation (CRG), 08003 Barcelona, Catalonia, Spain. ¹⁹CNRS, UMR 6291, 44007 Nantes, France. ²⁰Université de Nantes, 44322 Nantes, France. ²¹CHU Nantes, l'Institut du Thorax, Service de Cardiologie, 44093 Nantes, France. ²²Service d'Anatomie Pathologique, Hôpital Européen Georges Pompidou, 75015 Paris, France. ²³National Heart and Lung Institute, Harfield, Heart Science Centre, Imperial College London, London SW7 2AZ, UK. ²⁴Howard Hughes Medical Institute, Waksman Institute and Department of Molecular Biology and Biochemistry, Rutgers, the State University of New Jersey, Piscataway, New Jersey 08854, USA. ²⁵INSERM UMR_S910, Team physiopathology of cardiac development Aix-Marseille University, Medical School La Timone, 13885 Marseille, France. ²⁶Department of Cellular and Molecular Biology, University of Texas Health Science Center Northeast Tyler, Texas 75708, USA. ²⁷Gazes Cardiac Research Institute, Division of Cardiology, Department of Medicine, Medical University of South Carolina, Charleston, South Carolina 29425, USA. ²⁸Department of Radiology and Radiological Sciences, Medical University of South Carolina, Charleston, South Carolina 29425, USA. ²⁹Assistance Publique – Hôpitaux de Paris, Département de Génétique, Hôpital Européen Georges Pompidou, 75015 Paris, France. ³⁰Assistance Publique – Hôpitaux de Paris, Département de Cardiologie, Hôpital Européen Georges Pompidou, 75015 Paris, France. ³¹Cardiac Ultrasound Laboratory, Cardiology Division, Massachusetts General Hospital, Harvard Medical School, 55 Fruit Street, Boston, Massachusetts 02114, USA.

*These authors contributed equally to this work.

§These authors jointly supervised this work.

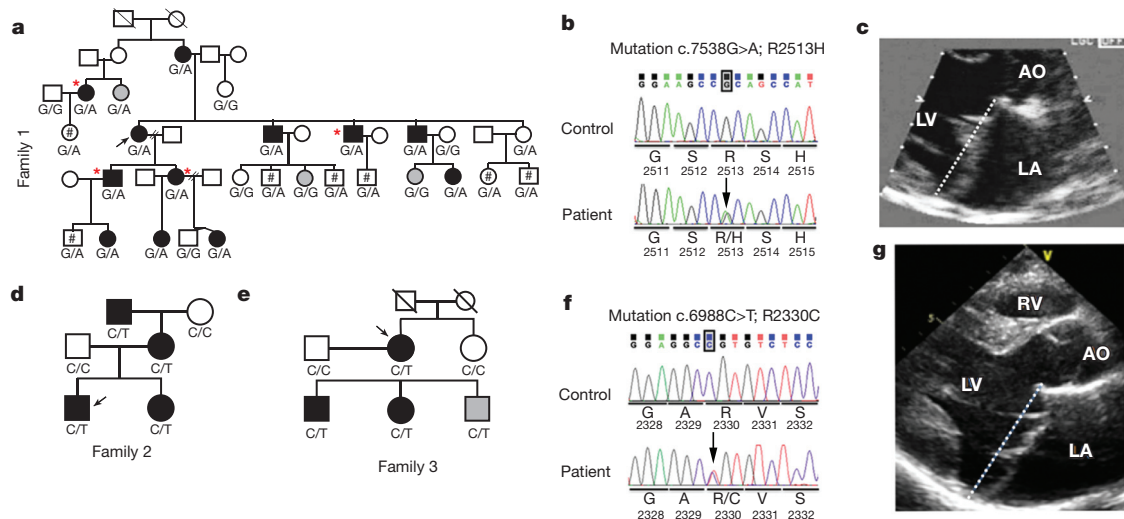


Figure 1 | Pedigrees, mutation, and phenotype. Black symbols, MVP affected; grey, unknown; arrows, probands. If no genotype is shown, the individuals were unavailable for study. **a**, Pedigree linked to chromosome 11. #, Individuals under 15 years of age; *, individuals sequenced. Genotypes c.7538G>A (R2513H) of *DCHS1* mutation are shown. **b**, DNA sequence of

c.7538G>A (p.R2513H). **c**, Two-dimensional echocardiographic long-axis view of family 1 proband. Dashed line marks mitral annulus. **d**, **e**, Family 2 and 3 pedigrees. Genotype c.6988C>T (p.R2330C) shown. **f**, DNA sequence c.6988C>T (p.R2330C). **g**, Two-dimensional echocardiographic long-axis view of family 2 proband.

(Extended Data Fig. 1a, b); however, knockdown of *dachsous1b* (*dchs1b*) led to significant changes in cardiac morphology (Fig. 2a; Extended Data Fig. 1a). Control zebrafish hearts undergo looping and develop an atrioventricular constriction by 48 h post-fertilization (hpf), whereas *dchs1b* knockdown disrupts this process, resulting in impaired formation of the atrioventricular constriction (Fig. 2a, b). While control embryos have unidirectional blood flow between the atrium and ventricle at 72 hpf (Supplementary Video 4), *dchs1b* knockdown causes regurgitation of blood from the ventricle into the atrium (Supplementary Video 5). An atrioventricular canal defect was defined as failure of cardiac looping combined with any atrioventricular regurgitation at 72 hpf. Using a high morpholino dose (1.5 ng) to establish the phenotype, the prevalence of atrioventricular canal defects was 76% ($n = 170$), whereas spontaneous cardiac defects were rarely observed in controls (0.5%, $n = 205$) (Fig. 2b). Whole-mount *in situ* hybridization of *dchs1b* confirmed predominant expression at the atrioventricular junction at 54 and 72 hpf, corresponding to the temporal defects observed in the morphants (Extended Data Fig. 3a–c). We evaluated gene expression patterns in the developing atrioventricular ring, and observed that *bmp4* expression is expanded into the ventricle at 48 hpf in *dchs1b* knockdown embryos while it is restricted to the atrioventricular ring in controls (Extended Data Fig. 4a, b). Additionally, *has2* expression was not detectable at 48 hpf, and only faintly at 72 hpf in the *dchs1b* knockdown (Extended Data Fig. 4i–l). To test mutation pathogenicity in this model, rescue experiments were performed using both wild-type human *DCHS1* and P197L/R2513H mutant mRNA, which were injected into *dchs1b* knockdown zebrafish with a lower dose of morpholino (0.75 ng) to minimize combined morpholino/mRNA toxicity. Human wild-type *DCHS1* mRNA rescued the atrioventricular canal defect observed upon *dchs1b* knockdown, whereas injection of an equimolar amount of mutant *DCHS1* mRNA failed to rescue (Fig. 2c). Injection of the mutant *DCHS1* mRNA alone did not cause atrioventricular canal defects, supporting a loss-of-function mechanism for the *DCHS1* mutation.

Having demonstrated segregation of a loss-of-function *DCHS1* mutation with MVP in our large pedigree, we sought to determine if genetic variation in *DCHS1* plays a role in MVP beyond the linked family. By evaluating a cohort of MVP patients, we identified two additional families in which MVP segregated with the novel *DCHS1* protein variant p.R2330C (Fig. 1d–g). The proband of family 2 underwent surgical mitral valve repair for severe MVP and mitral

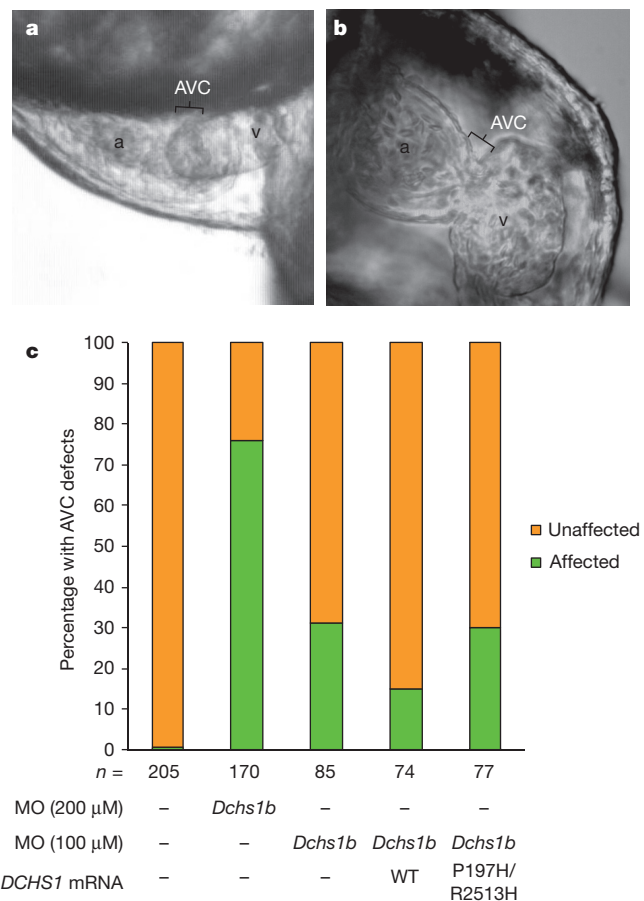


Figure 2 | Zebrafish *Dchs1b* is required for atrioventricular canal development. **a**, By 72 hpf, zebrafish hearts develop a constriction in the atrioventricular canal (AVC) that separates the atrium (a) from the ventricle (v). **b**, Knockdown of *dchs1b* results in absence of the atrioventricular constriction (bracket). **c**, Approximately 75% of *Dchs1b* morphants exhibit AVC defects ($*P = 1 \times 10^{-62}$). *DCHS1* human mRNA rescues the *dchs1b* morpholino AVC phenotype, whereas human mutant *DCHS1* mRNA (P197H/R2513H) fails to rescue the phenotype ($**P = 0.009$). The total number of fish analysed was 611 and statistical values were obtained using Fisher's exact test.

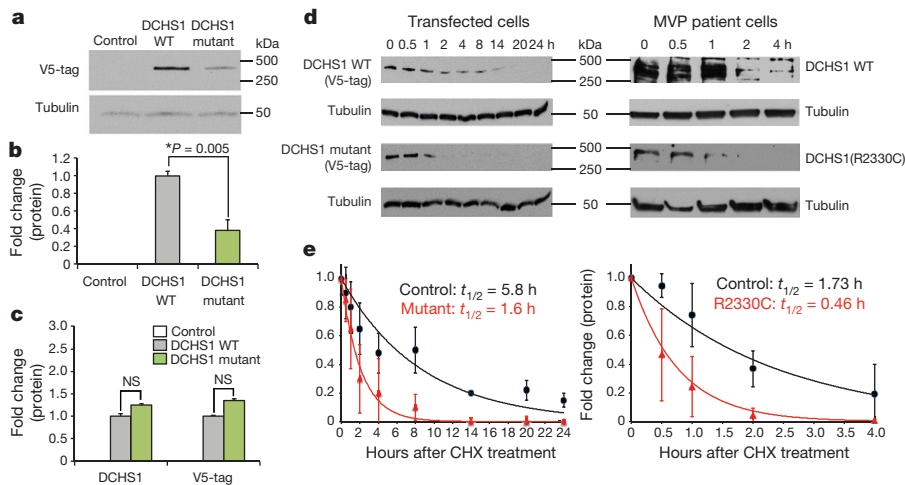


Figure 3 | *DCHS1* mutations result in diminished protein levels. **a–c**, Western blot, (p.P197L/p.R2513H) mutant *DCHS1* results in a 60% decrease in protein with no change in RNA expression. **d**, Left panel, *DCHS1* wild-type (WT) or mutant (p.P197L/p.R2513H) transfectants treated with cycloheximide (CHX) for specified times followed by western blot analyses. Right panel, cycloheximide on control (*DCHS1* WT) or MVP patient (p.R2330C) MVICs. Tubulin, loading control. **e**, Calculated protein half-lives. WT and mutant transfectants half-life = 5.8 h versus 1.6, respectively (left). Control and mutant *DCHS1* half-life is 1.73 h versus 0.46 h, respectively (right). Analyses performed in triplicate and repeated four times. Error bars, standard deviations; P values calculated using two-tailed Student's t -test.

regurgitation at age 21. Valve tissue was resected to repair the posterior leaflet and examination of the tissue showed classic myxomatous degeneration^{19,20} (Extended Data Fig. 5). Mitral valve interstitial cells were isolated from a portion of the posterior leaflet resected during surgery, providing a unique resource for the functional studies described below. The proband's sister, evaluated at age 27 and also heterozygous for the p.R2330C mutation, demonstrated classical MVP with thickened leaflets and moderate regurgitation. The father and maternal grandmother were unaffected and do not carry the mutation. The mother (age 49) and maternal grandfather (age 76) are both affected with mild MVP and both carry p.R2330C. In family 3, the proband (Fig. 1e) had moderate to severe heart failure owing to severe mitral regurgitation with posterior leaflet prolapse requiring surgery at age 72. The proband's sister, age 69, is unaffected and negative for the p.R2330C mutation. The son (age 52) and the daughter (age 53) are both affected with MVP and both carry p.R2330C. The second son (age 55) also carries p.R2330C, however his MVP status is indeterminate due to mild left ventricular inferior wall hypokinesia that tethers the leaflets down into the left ventricular cavity²¹, masking leaflet prolapse motion towards the left atrium.

To evaluate the functional consequence of the *DCHS1* mutations, we quantified protein levels in cells transfected with either wild-type or variant (p.P197L/p.R2513H) *DCHS1* complementary DNA constructs. Expression of the *DCHS1* mutant protein was ~60% less than

wild-type with no significant change in mRNA levels, suggesting that the *DCHS1* variants reduce protein stability (Fig. 3a–c). Cycloheximide treatment revealed that wild-type *DCHS1* protein in transfected cells had a half-life of 5.8 h, while the mutant *DCHS1* protein (p.P197L/p.R2513H) had a half-life of 1.6 h (Fig. 3d, e). *DCHS1* constructs harbouring either p.P197L or p.R2513H were evaluated, showing that p.R2513H markedly reduced protein levels, implicating this variant as pathogenic in the family (Fig. 1, Extended Data Fig. 6). A similar analysis of *DCHS1* protein half-life was conducted using p.R2330C MVICs from the proband of family 2. Consistent with the data obtained from the p.R2513H transfectants, these studies showed significant reduction in protein half-life compared to control MVICs ($t_{1/2} = 0.46$ h versus 1.73 h, Fig. 3d, e). Together, our studies show that p.R2513H and p.R2330C result in *DCHS1* loss of function. In order to evaluate *DCHS1* loss of function in a mammalian model, we analysed *Dchs1*-deficient mice for phenotypic similarities to human MVP.

Homozygous knockout of *Dchs1* in mice results in neonatal lethality and multi-organ impairment²²; however, the relevant genetic model for human MVP is the heterozygous *Dchs1* mouse. *Dchs1*^{+/-} mice exhibit mitral valve prolapse with pronounced involvement of the posterior leaflet, which is elongated and shifts the leaflet coaptation anteriorly (Fig. 4) as in the proband from family 1 (Fig. 1c, Supplementary Videos 1–3, 6, 7)²³. Micro-MRI analyses and 3D reconstructions of adult *Dchs1*^{+/-} mice reveal prominent posterior leaflet

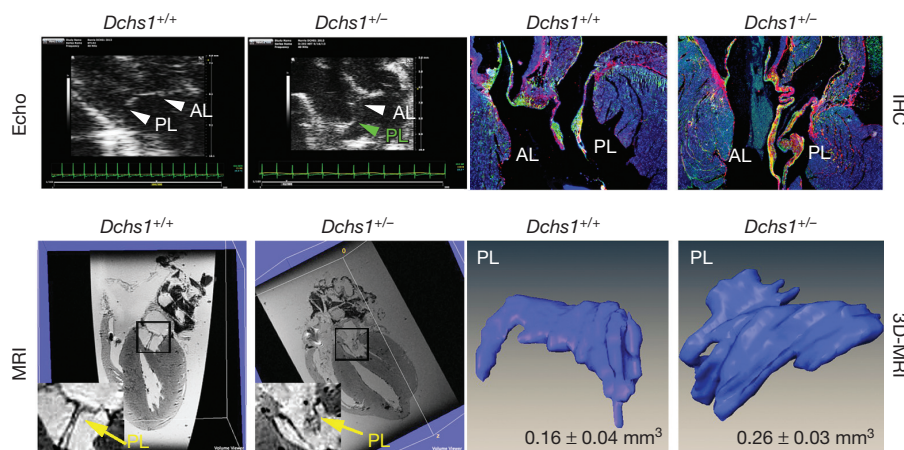


Figure 4 | *Dchs1* deficiency causes MVP and myxomatous degeneration in the adult mouse. Echocardiography (Echo), MRI, histopathology and 3D reconstructions performed on 9-month old male *Dchs1*^{+/+} and *Dchs1*^{+/-} mouse hearts. Echo, posterior leaflet prolapse in *Dchs1*^{+/-} (green arrow) ($n = 6$ per genotype). Immunohistochemistry (IHC), *Dchs1*^{+/-} ($n = 5$) anterior, posterior leaflets (AL, PL) exhibit myxomatous degeneration and expansion

of proteoglycan expression compared to *Dchs1*^{+/+} ($n = 7$), collagen (red), proteoglycans (green). MRI show posterior leaflet (PL) thickening in *Dchs1*^{+/-} (arrow-inset) compared to control littermates. 3D reconstructions of MRI: *Dchs1*^{+/-} mice exhibit thickened and elongated leaflets compared to *Dchs1*^{+/+}. (Two-tailed Student's t -test was used to calculate P values; $P = 0.01$, $n = 4$ per genotype).

thickening with quantitative increases in valve volume (Fig. 4 and Supplementary Video 8). All other echocardiographic measurements were unchanged (Extended Data Fig. 7). Histological and molecular characterization of *Dchs1*^{+/-} mice confirmed leaflet thickening and showed myxomatous degeneration with increased proteoglycan accumulation in both mitral leaflets (Fig. 4). These results clearly show that *Dchs1* heterozygosity results in mitral valve prolapse in mice.

To determine if MVP has a developmental origin, we performed expression and functional analyses at embryonic and fetal time points. RNA *in situ* hybridization and immunohistochemistry showed expression of *Dchs1* in endocardial and mesenchymal cells of atrioventricular valve leaflets at all time points examined (Extended Data Fig. 8). While no morphological defects were observed in the *Dchs1*^{+/-} mice during early embryonic development (E11.5–E13.5), at later time points (E15.5–E17.5) *Dchs1*^{+/-} mice displayed changes in mitral-valve shape (Fig. 5a–c), which were more severe in *Dchs1*^{-/-} animals. Histology and three-dimensional reconstructions of anterior and posterior mitral leaflets at E17.5 of *Dchs1*^{+/+}, *Dchs1*^{+/-} and *Dchs1*^{-/-} mice showed comparable leaflet volumes. However, *Dchs1*^{+/-} and *Dchs1*^{-/-} animals exhibited statistically significant changes in valve length and width (Supplementary Video 9 and Fig. 5a–c). In most leaflet regions measured, *Dchs1*^{+/-} animals displayed an intermediate phenotype, demonstrating a gene dosage effect. These shape changes implicate *Dchs1* as critical for proper anatomical patterning of the valve, consistent with previous reports of *dachsous* function in the *Drosophila* wing²⁴. Thus, *in vivo* lineage-tracing studies were performed on *Dchs1*^{+/+} and *Dchs1*^{+/-} mice. Crossing the *WT1-Cre/ROSA-eGFP*²⁵ line onto both *Dchs1*^{+/+} and *Dchs1*^{+/-} backgrounds allowed visualization of patterning defects of epicardial-derived cells (EPDCs) during migration into the posterior leaflet. This EPDC population initially migrates into the posterior leaflet as a sheet of cells. However, in the *Dchs1*^{+/-} mice this sheet-like appearance is disrupted and an increase in EPDCs infiltrating diffusely

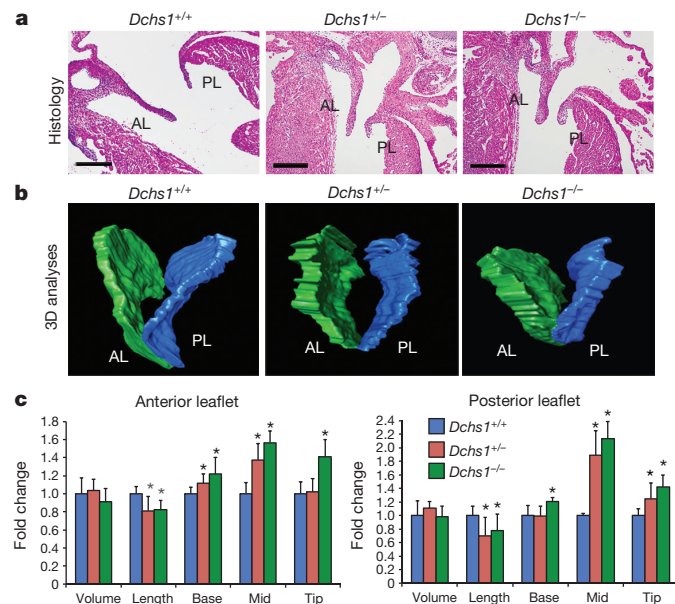


Figure 5 | Developmental aetiology for MVP. **a, b**, Haematoxylin and eosin and 3D reconstructions of E17.5 *Dchs1*^{+/+}, *Dchs1*^{+/-} and *Dchs1*^{-/-} mouse hearts showing thickening of anterior and posterior leaflets (AL, PL) in *Dchs1*^{-/-} mice compared to *Dchs1*^{+/+}. *Dchs1*^{+/-} valves display an intermediate phenotype. **c**, Quantification of valve dimensions showing *Dchs1*^{-/-} (green bars) and *Dchs1*^{+/-} (red bars) anterior and posterior lengths were significantly reduced compared to *Dchs1*^{+/+} (blue bars) leaflets. *Dchs1*^{-/-} and *Dchs1*^{+/-} valves displayed increased thickness throughout the leaflets compared to *Dchs1*^{+/+}. Scale bars, 100 μ m. *n* = 5 per genotype and two-tailed Student's *t*-test was used to calculate *P* values; **P* < 0.01.

throughout the valve tissue is observed (Extended Data Fig. 9a–c, Supplementary Video 10), concomitant with altered cellular patterning and alignment (Extended Data Fig. 9d, e). *In vitro* studies of MVICs from *Dchs1*^{+/-} and from the family 2 MVP proband (p.R2330C) also show this increased migration phenotype (Extended Data Fig. 10a, b). Taken together, the mouse and human studies support a developmental aetiology for MVP, and invoke a model for MVP in which cell migration and patterning defects mediated by *DCHS1* contribute to disease pathogenesis.

MVP is one of the most common cardiovascular diseases, affecting nearly 1 in 40 people worldwide^{1–3}. Although its heritability and variable expression in large pedigrees has been known for decades, its genetic underpinnings have remained elusive^{23,26}. We report the discovery of two loss-of-function mutations in *DCHS1* that segregate with MVP in three families. Our mouse models exhibit classical MVP, a phenotype that was traced back to developmental errors during valve morphogenesis. These findings provide a model for understanding inherited non-syndromic MVP as a developmentally based disease that progresses over the lifespan of affected individuals, consistent with previous reports on the natural history of MVP²⁷. A robust estimate of the total contribution of rare *DCHS1* genetic variation to sporadic MVP has yet to be determined and will require sequencing of large cohorts of MVP patients. Nonetheless, discovery of this novel mechanistic pathway elucidated by intensively studying rare familial mutations will facilitate the identification of additional MVP genes and reveal pathogenic mechanisms that hold the potential for pre-surgical therapy for this very common cardiac disease.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 23 December 2014; accepted 17 June 2015.

Published online 10 August 2015.

1. Freed, L. A. *et al.* Mitral valve prolapse in the general population: the benign nature of echocardiographic features in the Framingham Heart Study. *J. Am. Coll. Cardiol.* **40**, 1298–1304 (2002).
2. Avierinos, J. F. *et al.* Natural history of asymptomatic mitral valve prolapse in the community. *Circulation* **106**, 1355–1361 (2002).
3. Freed, L. A. *et al.* Prevalence and clinical outcome of mitral-valve prolapse. *N. Engl. J. Med.* **341**, 1–7 (1999).
4. Vaishnava, P., Fuster, V., Goldman, M. & Bonow, R. O. Surgery for asymptomatic degenerative aortic and mitral valve disease. *Nat. Rev. Cardiol.* **8**, 173–177 (2011).
5. Waller, B. F., Maron, B. J., Del Negro, A. A., Gottdiener, J. S. & Roberts, W. C. Frequency and significance of M-mode echocardiographic evidence of mitral valve prolapse in clinically isolated pure mitral regurgitation: analysis of 65 patients having mitral valve replacement. *Am. J. Cardiol.* **53**, 139–147 (1984).
6. Freed, L. A. *et al.* A locus for autosomal dominant mitral valve prolapse on chromosome 11p15.4. *Am. J. Hum. Genet.* **72**, 1551–1559 (2003).
7. Levine, R. A. *et al.* Three-dimensional echocardiographic reconstruction of the mitral valve, with implications for the diagnosis of mitral valve prolapse. *Circulation* **80**, 589–598 (1989).
8. Levine, R. A., Stathogiannis, E., Newell, J. B., Harrigan, P. & Weyman, A. E. Reconsideration of echocardiographic standards for mitral valve prolapse: lack of association between leaflet displacement isolated to the apical four chamber view and independent echocardiographic evidence of abnormality. *J. Am. Coll. Cardiol.* **11**, 1010–1019 (1988).
9. Perloff, J. K. & Child, J. S. Clinical and epidemiologic issues in mitral valve prolapse: overview and perspective. *Am. Heart J.* **113**, 1324–1332 (1987).
10. Clark, H. F. *et al.* *Dachsous* encodes a member of the cadherin superfamily that controls imaginal disc morphogenesis in *Drosophila*. *Genes Dev.* **9**, 1530–1542 (1995).
11. Adzhubei, I. A. *et al.* A method and server for predicting damaging missense mutations. *Nature Methods* **7**, 248–249 (2010).
12. Chun, S. & Fay, J. C. Identification of deleterious mutations within three human genomes. *Genome Res.* **19**, 1553–1561 (2009).
13. Schwarz, J. M., Rodelsperger, C., Schuelke, M. & Seelow, D. MutationTaster evaluates disease-causing potential of sequence alterations. *Nature Methods* **7**, 575–576 (2010).
14. Visel, A., Thaller, C. & Eichele, G. GenePaint.org: an atlas of gene expression patterns in the mouse embryo. *Nucleic Acids Res.* **32**, D552–D556 (2004).
15. Guénette, S. *et al.* Essential roles for the FE65 amyloid precursor protein-interacting proteins in brain development. *EMBO J.* **25**, 420–431 (2006).
16. Margolin, D. H. *et al.* Ataxia, dementia, and hypogonadotropism caused by disordered ubiquitination. *N. Engl. J. Med.* **368**, 1992–2003 (2013).
17. Golzio, C. *et al.* *KCTD13* is a major driver of mirrored neuroanatomical phenotypes of the 16p11.2 copy number variant. *Nature* **485**, 363–367 (2012).

18. Chaki, M. *et al.* Exome capture reveals *ZNF423* and *CEP164* mutations, linking renal ciliopathies to DNA damage response signaling. *Cell* **150**, 533–548 (2012).
19. Norris, R. A. *et al.* Expression of the familial cardiac valvular dystrophy gene, *filamin-A*, during heart morphogenesis. *Dev. Dyn.* **239**, 2118–2127 (2010).
20. Rabkin, E. *et al.* Activated interstitial myofibroblasts express catabolic enzymes and mediate matrix remodeling in myxomatous heart valves. *Circulation* **104**, 2525–2532 (2001).
21. Nesta, F. *et al.* Leaflet concavity: a rapid visual clue to the presence and mechanism of functional mitral regurgitation. *J. Am. Soc. Echocardiogr.* **16**, 1301–1308 (2003).
22. Mao, Y. *et al.* Characterization of a *Dchs1* mutant mouse reveals requirements for *Dchs1*-*Fat4* signaling during mammalian development. *Development* **138**, 947–957 (2011).
23. Nesta, F. *et al.* New locus for autosomal dominant mitral valve prolapse on chromosome 13: clinical insights from genetic studies. *Circulation* **112**, 2022–2030 (2005).
24. Cho, E. & Irvine, K. D. Action of *fat*, *four-jointed*, *dachsous* and *dachs* in distal-to-proximal wing signaling. *Development* **131**, 4489–4500 (2004).
25. Wessels, A. *et al.* Epicardially-derived fibroblasts and their contribution to the parietal leaflets of the atrioventricular valves in the murine heart. *Dev. Biol.* **366**, 111–124 (2012).
26. Zuppiroli, A., Roman, M. J., O'Grady, M. & Devereux, R. B. A family study of anterior mitral leaflet thickness and mitral valve prolapse. *Am. J. Cardiol.* **82**, 823–826 A10 (1998).
27. Kolibash, A. J. Jr *et al.* Evidence for progression from mild to severe mitral regurgitation in mitral valve prolapse. *Am. J. Cardiol.* **58**, 762–767 (1986).

Supplementary Information is available in the online version of the paper.

Acknowledgements This work was supported by the Fondation Leducq (Paris, France) Mitral Transatlantic Network of Excellence grant 07CVD04. MVP patient studies were supported by an Innovation in Clinical Research award of the Doris Duke Charitable Foundation, by an award of the Aetna Quality Care Research Fund, and by a gift from Rena M. Shulsky, New York, New York (S.A.S. and R.A.L.). Sequencing of the candidate region was performed at the Venter Institute through a grant from the National Heart Lung and Blood Institute Resequencing and Genotyping (RS&G) Service (S.A.S.). The work at MUSC was performed in a facility constructed with support from the National Institutes of Health, Grant Number C06 RR018823 from the Extramural

Research Facilities Program of the National Center for Research Resources. Collection of the MVP France cohort was supported by the French Society of Cardiology. Other funding sources: National Heart Lung and Blood Institute: R01HL122906-01 (A.W.), R01-HL33756 (R.R.M.), COBRE 1P30 GM103342 (R.R.M., R.A.N., A.W.), 8P20 GM103444-07 (R.R.M. and R.A.N.), R01-HL109004 (D.J.M.), R01-HL127692 (D.J.M., S.A.S., R.A.N., R.A.L.); R01-HL095696 (D.R.M.), VA Merit Review BX002327 (D.R.M.); National Institute of Mental Health R00-MH095867 (M.E.T.) The Hassenfeld Scholar Program (D.J.M.); The March of Dimes (M.E.T.); M.G.H. Scholars Program (S.A.S., M.E.T.); American Heart Association: 09GRNT2060075 (A.W.), 11SDG5270006 (R.A.N.), 2261354 (D.J.M.), 15GRNT25080052 (R.A.N.); National Science Foundation: EPS-0903795 (R.R.M.); NHLBI K24 HL67434, R01HL72265 and R01HL109506 and the Ellison Foundation, Boston, MA (R.A.L.), Howard Hughes Medical Institute (K.D.I.), and a gift from Michael Zak (D.J.M.). Thanks to T. Brown (MUSC) for his guidance on MRI studies, C. Hanscom (M.G.H.) for assistance with genomic libraries, and E. Lim (H.M.S.) for contributions in interpreting the exome mutation data. The authors would like to thank the Exome Aggregation Consortium and the groups that provided exome variant data for comparison. A full list of contributing groups can be found at <http://exac.broadinstitute.org/about>.

Author Contributions R.D., M.L., C.S., C.J., M.P., X.J., J.-J.S., D.T., S.O., X.E., F.C., and S.A.S. participated in genetic analysis, sequencing and mutation cloning. R.D., F.N.D., L.A.F., T.L.T., H.L.M., L.F.-F., J.S., C.T., R.A.L., and A.H. participated in patient collection and phenotyping using echocardiography. D.S.P., S.N.L. and D.J.M. performed zebrafish and cell culture experiments. M.E.T., M.R.S., N.B.N., C.D., H.B. and C.C. performed bioinformatics and statistical analysis. A.C., P.C., and P.B. established human patient cell cultures and histology on human tissues. A.d.V., K.W., K.D.I., Y.M., K.S., A.W., T.M., K.T., R.R.M. and R.A.N. performed mouse embryo and knockout experiments and cell alignment and migration assays. X.N., A.-M.B., D.R.M., H.K. performed mouse *in vivo* imaging. R.D., D.J.M., R.A.L., R.A.N. and S.A.S. wrote the manuscript. R.A.L., A.H., S.A.S. and J.J.S. coordinated the Leducq Mitral Network.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.A.S. (slaugenhaupt@chgr.mgh.harvard.edu) or R.A.N. (norrisra@musc.edu).

METHODS

Study participants. Family 1 was originally recruited through the Echocardiography Laboratory at Massachusetts General Hospital as part of a phenotype-driven genetic study of MVP. MVP was diagnosed by specific criteria (>2 mm atrial leaflet displacement in a parasternal long-axis view)^{6–9}. The study was approved by the Institutional Review Board of Partners Healthcare, Boston, Massachusetts, and all participants provided written informed consent. Complete details of the linkage analysis on this large, multigenerational family have been previously published⁶. In brief, the family contains 41 individuals in five generations. Echocardiograms and DNA were obtained on 28 subjects, of whom 12 were diagnosed with MVP, three were classified as having nondiagnostic minimal leaflet displacement, and 13 were unaffected⁶. Three patients had non-diagnostic valve leaflet displacement and were considered unknown for the original linkage analysis. The proband had prominent MVP with thickened leaflets, severe mitral regurgitation, and heart failure ultimately requiring surgical valve repair (Fig. 1c and Supplementary Videos 1–3). Other affected members also showed diffuse leaflet thickening, prolapse, and mitral regurgitation of varying severity; one required surgical repair. No extracardiac manifestations of connective-tissue abnormalities or Marfan syndrome were present in any family members. Following a complete genome scan, parametric and non-parametric analyses confirmed linkage of this family to a 4.3 cM region of chromosome 11p15.4. Consistent with the model of sex- and age-dependent penetrance, several of the unaffected members who carried the MVP allele were less than 15 years old at the time of evaluation (Fig. 1a). Importantly, an analysis using only affected individuals confirmed the linkage result.

DNA sequencing and variant calling in family 1. In order to identify the mutation, four affected individuals who shared the disease haplotype were chosen for sequencing (Fig. 1a). To reduce the likelihood of random haplotype sharing, we selected individuals with four distinct haplotypes on the non-MVP allele. A 2.1 Mb region of human chromosome 11 (5094774–7248926; NCBI36 coordinates) was targeted and screened for repetitive regions using the SureSelect system (Agilent). DNA extraction was performed using the AutoGenFlex STAR automated system (Autogen) and FlexiGen DNA purification reagents (Qiagen) according to the manufacturers' instructions. Bait oligonucleotides were designed to the non-repetitive regions of the targeted linkage peak, resulting in 1.03 Mb of target sequence using the SureSelect in-solution long RNA baits (Agilent). Captured DNA was amplified and quantified using the Agilent High Sensitivity DNA Kit for the Agilent 2100 Bioanalyzer, and sequenced using Illumina sequencing chemistry (paired-end, 100 cycles) at the Venter Institute supported by the NHLBI Resequencing and Genotyping Program. One hundred bases were sequenced from each end of the captured DNA fragments. Image analysis and base calling were performed using Illumina's GA Pipeline version 1.6.0. Sequence reads were mapped to the human genome (ncbi36) and variants identified using clc-ngs-cell-2.0.5-linux_64 (clc_ref_assemble_long -q -p fb ss 180 360 -I -r, and find_variations -c 8 -v -f 0.2). Variants were classified using VariantClassifier²⁸. 4,891 SNVs were identified in the four subjects, 1,951 were shared by all four subjects. We classified all rare SNVs or those of unknown frequency based on conservation data, population genetic data, and predictive functional impact from public resources. We performed analyses of conservation of the variant locus using PhyloP²⁹, PhastCONS³⁰ and GERP³¹, and assessed the population frequency initially using the Exome Variant Server, NHLBI GO Exome Sequencing Project (ESP), Seattle, Washington (<http://evs.gs.washington.edu/EVS/>) as an initial filter and the 1000 Genomes Project³² as a secondary filter, cognizant of the limitations of the low-depth coverage of the 1000 Genomes Project to characterize rare mutations, which were available during our initial analyses.

Sporadic MVP patient cohort. As part of an international consortium on mitral valve disease, we initiated collection of MVP patients with the eventual goal of performing GWAS studies in MVP. To date, 1,896 patients have been collected in the United States, France, and Spain. MVP was defined as systolic displacement of one or both mitral leaflets ≥ 2 mm beyond the annulus in parasternal or apical long-axis views, asymmetric posterior leaflet prolapse was also included in any view, including apical 4-chamber, when confirmed by side-to-side long-axis scanning^{7,8}. Patients were required to have no evidence by history, physical examination, or imaging for Marfan syndrome or other connective tissue disorders associated with MVP.

Exome sequencing, genotyping, and variant evaluation. As part of an MVP exome sequencing pilot project conducted by the Leducq Mitral Network, exome data were generated on twenty-one severe, early onset MVP patients and made available to identify variants in *DCHS1*. Fifteen patients were collected in Paris and had severe bileaflet mitral valve prolapse with myxomatous leaflets and an average age of onset of 15 years. Six patients were collected in Nantes, France, and had similar clinical characteristics with an average age at onset of 42. Exome capture was carried out using the SureSelect Human All Exon System using the

manufacturer's protocol version 1.0 that is compatible with Illumina paired-end sequencing. Exome-enriched genomes were multiplexed by flow cell for 101-bp paired-end read sequencing according to the protocol for the HiSeq 2000 sequencer (version 1.7.0; Illumina) to allow a minimum coverage of 30 \times . Reads were aligned to the human reference genome (UCSC NCBI36/hg19) using the Burrows-Wheeler Aligner (version 0.5.9). Evaluation of the *DCHS1* gene yielded 4 novel coding sequence variants that confirmed following repeat Sanger sequencing: 6646587 G/A (p.R2330C) 6646709 G/A (p.A2289V), 6648584 C/T (p.A1896T), 6648820 A/G (p.V1817A), base pair positions are NCBI36 coordinates. These four variants, in addition to the variants identified in family 1, were genotyped using Sequenom technology in the sporadic cohort. The major steps included primer and multiplex assay design using Sequenom's MassARRAY Designer software, DNA amplification by PCR, post-PCR nucleotide deactivation using shrimp alkaline phosphatase (SAP) to remove phosphate groups from unincorporated dNTPs, single-base extension reaction for allele differentiation, salt removal using ion-exchange resin, and mass correlated genotype calling using SpectroCHIP array and MALDI-TOF mass spectrometry. Quality control to determine sample and genotyping quality and to potentially remove poor SNPs and/or samples was performed in PLINK, a whole genome association analysis toolset. We predicted the impact on gene function using PolyPhen²¹, Mutation Taster¹³ and LRT¹².

Identification of family 2 and 3. In order to identify other mutations in *DCHS1*, we first evaluated *DCHS1* in the exome sequence data described above, reasoning that early onset forms may be more likely to have strong genetic aetiologies. Rare variants causing amino acid substitutions in *DCHS1* were identified in four individuals (p.V1817A, p.A1896T, p.A2289V, and p.R2330C) and genotyped in a cohort of 1,864 sporadic MVP patients that included the 21 individuals with exome data; two of these variants, both localized to exon 19, were observed in the MVP cohort (p.A2289V in two cases and p.R2330C in three cases). The proband in family 2 carried the p.R2330C variant and underwent surgery for MVP in Paris. We were able to collect DNA and echocardiograms on first-degree relatives at that time. Additional clinical characteristics of the proband in family 2 included congestive heart failure (NYHA II/III) with left ventricular dilatation (70/50 mm end-diastolic/end-systolic dimensions), impaired left ventricular systolic function (ejection fraction 53%, low for this volume overload), recurrent symptomatic atrial fibrillation, non-sustained ventricular tachycardia, and exercise-induced pulmonary hypertension (70 mm Hg systolic). The proband in family 3 was originally collected in Amiens, France. All echocardiograms were read in both Boston and Paris and readers were blind to genotype data.

D. rerio studies. Husbandry, knockdown and expression analyses were performed in the wild-type *D. rerio* (zebrafish) strain Tubingen AB. Morpholinos were injected at a dose of 1.5 ng (after dose optimization) into single-cell embryos to achieve gene knockdown, and phenotypes were examined at 48 and 72 h post-fertilization in three separate experiments of 50–75 embryos and compared to controls using Fisher's exact test. Morpholino GeneTools LLC (Philomath, OR) sequences were as follows: *apbb1* AACAAAGCGTACCACTCAGATTAGC, *dchs1a* TAAAGAAATGACAGTCCTACCTCCA, and *dchs1b* CATAACTGTTAAGAGTTCGCTACA. Knockdown was confirmed by quantitative polymerase chain reaction. qPCR was performed as previously described³³. In brief, 20–30 morpholino-injected embryos were collected at 72 hpf, and snap frozen in liquid nitrogen. TRIzol (Sigma) was added, RNA was purified according to the manufacturer's instructions, and cDNA was prepared using a Superscript III Kit (Invitrogen). Primer sets were as follows: *apbb1*, 5'-GTGGAGCGAGAACAGAG, 5'-CCAGCAGGAAGATCCGTGTC; *dchs1a*, 5'-GTTTCATGGAGGTTACAGC, 5'-CTTAATCCACCCCATCCAC; *dchs1b*, 5'-GTTTCCTTGAGGTAAAGGCGG, 5'-GGCCACCCCATCGGACG. qPCR was performed using SYBR Green (Applied Biosystems) in triplicate on an Applied Biosystems 7500 Fast Real-Time PCR instrument and normalized against β -actin. All zebrafish experiments were performed under protocols approved by the Institutional Animal Care and Use Committee at Massachusetts General Hospital.

D. rerio in situ hybridizations. *In situ* hybridizations were performed as previously described³⁴ using a partial clone of *dchs1b* (Open Biosystems, Clone ID 7136458) amplified with primers containing a T7 RNA polymerase site engineered onto the 3' end of the reverse primer (Forward 5'-GGCAGTTCAGGTGGTGGT. Reverse: TAATACGACTCACTATAGGGTTAAATCCTCATCTCAGCCTCA, T7 site underlined.) The *dchs1b* probe was produced using a T7 RNA polymerase (Ambion) and digoxigenin-labelled dNTPs (Roche). Other riboprobes used in the study have been previously described³⁵.

Generation of DCHS1 expression constructs. Human *DCHS1* and the mutant containing the c.590C>T and c.7538G>A sequence changes were synthesized by Integrated DNA Technologies. A unique EcoRI site, a T7 polymerase site, and a Kozak sequence were added to the 5' end of each gene, while a V5 tag and unique XhoI site were added to the 3' end. Each gene was then subcloned into the expression vector pcDNA3.1. Additional expression constructs were generated

that contained only the c.590C>T (p.P197L) mutation or only the c.7538G>A (p.R2513H) mutation. These constructs were made using the QuikChange II XL Site-Directed Mutagenesis Kit (Agilent Technologies) as per the manufacturer's instructions. The P197L construct was generated from the double mutant construct by changing the R2513H (c.7538G>A) mutation back into the wild type sequence using the following primers: 5'-gctgatggaagcgcagccatgccgct, 3'-agcggc atggctggcgtccatcagc. (The underlined bold base indicates the base pair changed.) The R2513H construct was generated by introducing the (c.7538G>A) mutation into the wild type *DCHS1* construct using the following primers: 5'-gctgatggaagc acagccatgccgct, 3'-agcggc atggctggcgtccatcagc.

Preparation and injection of *DCHS1* mRNA. mRNA was prepared from the wild-type *DCHS1* and *DCHS1* mutant expression vectors using a T7 mMessage mMachine Kit (Ambion) according to the manufacturer's instructions. Injection mixtures containing 0.75 ng *dchs1b* MO alone, or 0.75 ng *dchs1b* MO plus 7 fg μl^{-1} of human *DCHS1* mRNA (either wild type or mutant) were injected into one-cell embryos, and fish were scored for atrioventricular canal defects (failure to loop, and presence of regurgitation) 72 h later. Data were collected from three independent experiments performed with 20–30 embryos each and comparisons made using Fisher's Exact test.

Isolation of *DCHS1* p.R2330C MVP and control patient mitral valve tissue and valvular interstitial cells. Resected posterior mitral valve tissue was used for culture and histology. For culture, valve pieces were minced in phosphate buffered saline (PBS) and washed in DMEM with antibiotics (penicillin/streptomycin (P/S) and fungizone) and incubated in DMEM with collagenase type II (Worthington) (1 mg ml^{-1}) at 37 °C for 12 h. Following mechanical dissociation in DMEM, the cell suspension was filtered through a 40- μm cell strainer and cells were cultured in DMEM with 15% fetal calf serum and antibiotics (P/S, fungizone). Although rare valve endothelial cells were present at P0, only cells with a fibroblastic phenotype (VICs) remained following P1–2. For all experiments, these valvular interstitial cells were used before passage 5. For histology: valves were fixed in formalin, embedded in paraffin and sectioned at 5 μm . Movat's Pentachrome histological stain was performed using standard procedures.

Cell culture studies. Wild type, p.P197L, p.R2513H and p.P197L/R2513H *DCHS1* constructs were either synthesized by Integrated DNA Technologies or generated by site-directed mutagenesis (as described above), with an amino-terminal V5 epitope tag. Except where indicated, "mutant *DCHS1*" indicates the double mutant p.P197L/p.R2513H haplotype in family 1. These constructs were expressed in mycoplasma-free HEK293 cells (ATCC, not independently authenticated) using cationic lipid-mediated transient transfection (Lipofectamine LTX, Invitrogen). Protein expression of transfected HEK cells was measured by quantifying western blots using an antibody to the V5 epitope tag (Invitrogen). Patient cells: for patient cells, control and p.R2330C valvular interstitial fibroblasts from posterior leaflets were plated at 2.5×10^4 cells in a 24-well dish. 24 h later, protein stability experiments were performed. Protein stability experiments involved addition of cycloheximide 24 h after transfection (WT and p.R2513H transfectants) or plating (control or p.R2330C patient cells). For the cycloheximide experiments, media containing 100 ng ml^{-1} of cycloheximide was added at 24 h post-transfection and each well was harvested as above at the indicated time points. Western blots were probed with either a mouse anti-V5 primary antibody (1:4,000 dilution Invitrogen) or a rabbit anti-*Dchs1* antibody³⁶ (1:1,500 dilution), and an HRP-linked secondary at the same dilution (Thermo Scientific). Blots were also probed with a mouse anti-tubulin primary (Millipore) at a 1:4,000 dilution, and the same secondary antibody as above. Blots were treated with Pierce ECL Substrate and visualized on film. For quantitation, blot pixel intensity was measured by ImageJ (NIH), and normalized to tubulin. Each sample was run in triplicate, and a regression curve was fit and half-lives calculated using SigmaPlot 12.

Mouse studies. *Dchs1* mice and genotyping were previously described²². All mice were blinded for genotype. Following phenotypic analyses, the genotypes of each sample were matched with the experimentally determined data sets. Histology, mitral valve analyses (echocardiography, MRI, and morphometric determination), and expression studies were performed on embryonic and adult (9-month male) wild-type (*Dchs1*^{+/+}), heterozygote (*Dchs1*^{+/-}), and knockout (*Dchs1*^{-/-}) hearts (C57/BL6;Sv129 mixed background). For histology: fetal (E17.5) and adult (9-month) hearts were processed for haematoxylin and eosin stainings and immunohistochemistry (IHC) as previously described³⁷. For fetal analyses, *Dchs1*^{+/+}, *Dchs1*^{+/-}, and *Dchs1*^{-/-} mice were analysed ($n = 5$ per genotype). Due to neonatal lethality of the *Dchs1*^{-/-} mice and loss-of-function *Dchs1* mutations in humans, adult analyses were restricted to *Dchs1*^{+/+} ($n = 7$) and *Dchs1*^{+/-} ($n = 5$). For all analyses male mice were used. Antibodies used for IHC were: hyaluronan binding protein (HABP) to stain proteoglycans (1:100) (Calbiochem), collagen I (1:100) (MDbio), and Hoechst to stain nuclei (1:10,000) (Invitrogen). AMIRA 3D reconstructions were performed to generate volumetric measurements of fetal (E17.5) anterior and posterior mitral leaflets ($n = 5$ for each genotype) as previously

described³⁸. Length and width measurements of the mitral leaflets were obtained from histological sections. 25 consecutive 5 μm sections from anterior and posterior mitral leaflets of each genotype were used for measurements (*Dchs1*^{+/+}, $n = 4$, *Dchs1*^{+/-}, $n = 7$, *Dchs1*^{-/-}, $n = 5$). ImageJ software was used to measure the length of anterior and posterior leaflets from annulus to tip and the perpendicular width at the base, mid-region and tip. Measurements were compared to wild-type data to generate fold change and statistical significance ($P < 0.01$) was calculated using a Student's *t*-test. For quantification of mitral valve interstitial cell alignment: hearts were initially dissected from E17.5 fetuses. The apex of the heart was dissected and discarded followed by removal of the left atrium. A cut was made cranial to caudal along the anterior aspect of the myocardium. The left ventricle was reflected to gain visualization of the leaflets. The left ventricle and interventricular septum were pinned such to stretch the papillary muscle and the chords. This resulted in obtainment of the leaflet as a planar sheet of tissue. The tissue was fixed in this position to ensure the leaflet was maintained in this orientation, as failure to do so results in curling of the leaflet making measurements and plane of orientation inconsistent between animals. Once the planar leaflet tissue was fixed in 4% PFA for 5 min, the leaflet was released from the heart by cutting the chords and dissecting along the annulus fibrosae. The tissue was then processed through normal protocols and placed en-face on paraffin. This technique was performed blinded to genotype and performed in exactly the same manner for all valve isolates. Performing this type of dissection and tissue processing ensures that all leaflets are placed in nearly identically oriented planes. The vector maps and quantification of cell alignment were performed blinded by two independent researchers. Only after the data was generated did a third researcher perform the PCR genotyping. Vector maps were manually generated as previously described³⁹. Cells that deviated >10 degrees from an average alignment plane were counted as misaligned. Interstitial cells within anterior leaflets from each genotype were measured. Total number of cells measured were: WT = 1,083, ($n = 4$); Het = 1,118, ($n = 4$); KO = 1,953, ($n = 4$). Statistical significance was calculated using a Student's *t*-test with a *P* value < 0.05 being significant. Very little variation existed between the independent valves of each genotype as is graphically depicted. For mouse echocardiography the Vevo2100 imaging system (VisualSonics, Toronto, Canada) was used with 22–55 MHz linear transducer probe (MS550D) and used for 2-D B-mode and M-mode analysis. Heart rate was maintained at 400–500 bpm via isoflurane anaesthesia. The mitral valve leaflet was visualized and its function was assessed in parasternal long-axis B-mode view by placing the transducer on the left lateral chest wall. End-systolic and end-diastolic left ventricular dimensions and wall thicknesses were measured according to the American Society of Echocardiography guidelines as applied to mice. Left ventricular wall thickness was measured at the level of interventricular septum and the posterior wall. Left ventricular volume was calculated from Simpson's method of disks and ejection fraction determined from the formula (left ventricular end-diastolic-end-systolic volume)/(left ventricular end-diastolic volume). Offline image analyses were performed using dedicated VisualSonics Vevo2100 1.2.0 software. Mitral valve prolapse was determined based on superior systolic displacement of one or more leaflets above the line connecting the annular hinge points in the long-axis view ($n = 6$ per genotype). For MRI experiments: 9-month-old male *Dchs1*^{+/+} and *Dchs1*^{+/-} mice ($n = 4$) were sacrificed and the hearts were perfusion fixed and immersed 1:40 (12.5 mmol) Gadolinium (ProHance) in 10% formalin overnight before imaging. MRI was undertaken at 7T using a Bruker Biospin console (Pavavision 5.1) with a volume transmitter coil and a phased array surface coil. Gradient echo FLASH 3D images were collected with repetition time/echo time = 50 ms/5.4 ms, flip angle = 30°, number of excitations = 3, matrix = 256 × 256 × 256 and pixel resolution = 55 × 55 × 59 μm . Images in DICOM format were imported into AMIRA 3D reconstruction software and volume quantification were performed as described previously³⁸. Pairwise comparison of littermates were performed and statistical significance was determined (Student's *t*-test) with $P < 0.01$. All mouse experiments were performed under protocols approved by the Institutional Animal Care and Use Committee, Medical University of South Carolina. Prior to cardiac resection, mice were euthanized in accordance with the Guide for the Care and Use of Laboratory Animals (NIH Publication No. 85-23, revised 1996).

RNA expression analyses of *Dchs1* and *Apbb1*. Section *in situ* hybridization was performed as previously described⁴⁰ on 4 embryos at each time point to localize *Dchs1* expressing cells throughout cardiac development. A *Dchs1* digoxigenin-labelled riboprobe (Roche) was generated against region 9222–10180 of accession number NM_00162943 and used for *in situ* hybridization at E11.5, E13.5, and E15.5. RNA *in situ* hybridization for *Apbb1* at E14.5 was performed through GenePaint¹⁴. Two separate riboprobes were used to analyse *Apbb1* RNA expression at E14.5. These probes were generated against regions 1676–2506 and 370–1967 of accession number NM_001253885.1. These probes span all known isoforms for *Apbb1* and provide similar spatial RNA expression patterns.

Protein expression. *Dchs1* antibodies were generated by immunizing rabbits with a synthetic peptide corresponding to rat *Dchs1* protein sequence: CSTYMVES PDLVEADSAA (region 1308–1324 of accession number NP_001101014)³⁶. Immunohistochemistry was performed as described previously¹⁹ using a 1:100 dilution of primary antibody.

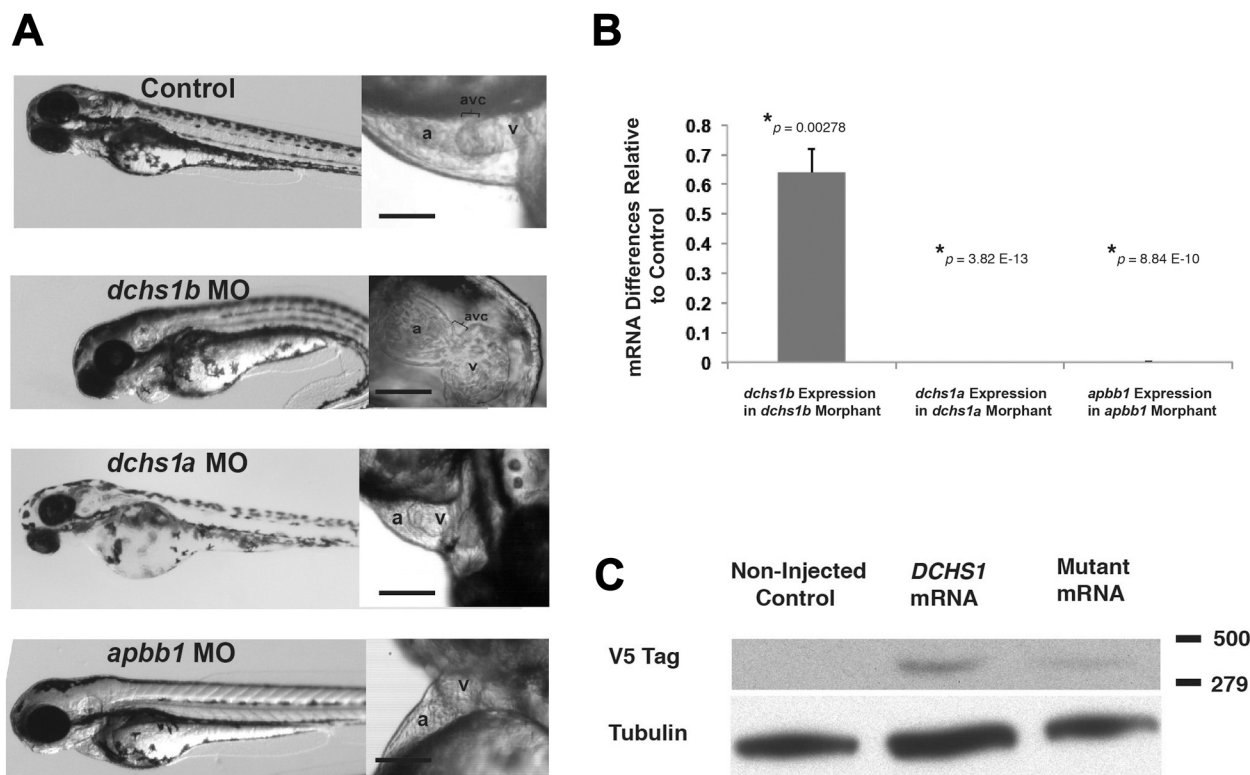
In vivo lineage trace. To trace the fate of epicardially derived cells in *Dchs1*^{+/+} and *Dchs1*^{+/-} mitral leaflets, the Wt1/IRES/GFP-Cre mouse²⁵ was bred with the *Dchs1*^{+/-} and *Dchs1*^{+/+} mice ($n = 4$ per genotype). Mice were euthanized at neonatal day 0 (P0), hearts were isolated and fixed overnight at 4 °C in 4% paraformaldehyde dissolved in PBS. Hearts were processed through a series of graded ethanol, cleared in toluene, and embedded in Paraplast Plus (Fisherbrand, 23-021-400). Hearts were sectioned at 5 µm and slides were treated with 15 ml of antigen unmasking solution (Vector Biolabs, H-3300) in 1,600 ml of distilled water for 10 min in a pressure cooker (Cuisinart) followed by incubation for 1 h at room temperature with 1% BSA (Sigma, B4287) in PBS. Expression of EGFP after Cre recombination was detected by immunofluorescence using antibodies against GFP (Abcam, 13970) and myosin heavy chain (MF20; DSHB). 5 µm sections throughout the entire valve were used for 3D reconstructions using Amira software. The volume of GFP positive cells and the volume of each mitral leaflet were measured using this software. Cell counting was done on GFP positive and GFP negative cells every 15 µm throughout the entire valve. Pairwise comparison of littermates were performed and statistical significance was determined (Student's *t*-test) with $P = 0.04$ for posterior leaflet and $P = 0.86$ for anterior leaflet.

In vitro migration. Human mitral valve interstitial cells were isolated from a control and the patient with the *DCHS1* mutation (p.R2230C) (proband family 2) and seeded into the Radius 24-well Cell Migration Assay plate containing hydrogels (Cell Biolabs, CBA-125). Cells were allowed to adhere overnight and then gels were dissolved. Wells were imaged over a period of 24 h and area of the cell free region was measured in Photoshop v.10.0.01 and subtracted from the initial area of the hydrogel to generate area migrated over time. Migration in the *Dchs1*^{+/-} and *Dchs1*^{+/+} mice was assessed by explanting P0 neonatal posterior mitral leaflets onto plastic. Images of the explants and migrating cells were captured at multiple time points. Distance migrated was measured as the distance from the explant to the farthest migrating cell. Measurements were taken at 5 points around the explant and averaged to calculate distance migrated. After 24 h, cells were fixed in ice-cold 100% methanol for 10 min and immunofluorescence was performed using an antibody against N-cadherin (1:1,000 dilution, BD

Transduction Labs, 610920). Pairwise comparisons were performed and statistical significance was determined (Student's *t*-test) with $P < 0.05$.

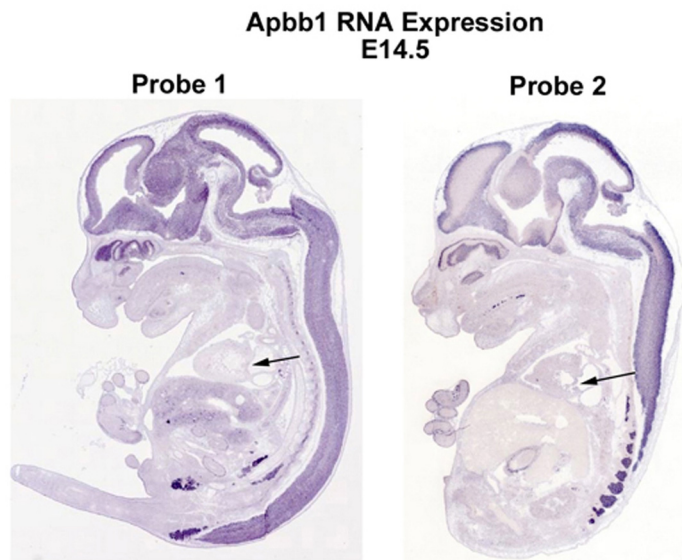
Statistical considerations. In all experiments sample sizes were chosen to provide power of 0.8 to detect biologically significant differences between test groups with two-sided $\alpha = 0.05$. Specific statistical tests are listed in the methods for each individual experiment. Assumptions of normal distributions were made for quantitative biological measurements and comparison groups were assumed to have similar variances. For zebrafish experiments, fertilized oocytes were randomly selected within each clutch for injection with active compound versus controls. Mouse experiments were interpreted blinded to genotype.

28. Li, K. & Stockwell, T. B. VariantClassifier: a hierarchical variant classifier for annotated genomes. *BMC Res. Notes* **3**, 191 (2010).
29. Pollard, K. S., Hubisz, M. J., Rosenbloom, K. R. & Siepel, A. Detection of nonneutral substitution rates on mammalian phylogenies. *Genome Res.* **20**, 110–121 (2010).
30. Siepel, A. *et al.* Evolutionarily conserved elements in vertebrate, insect, worm, and yeast genomes. *Genome Res.* **15**, 1034–1050 (2005).
31. Cooper, G. M. *et al.* Distribution and intensity of constraint in mammalian genomic sequence. *Genome Res.* **15**, 901–913 (2005).
32. The 1000 Genomes Project Consortium. An integrated map of genetic variation from 1,092 human genomes. *Nature* **491**, 56–65 (2012).
33. Morrison, T. B., Weis, J. J. & Wittwer, C. T. Quantification of low-copy transcripts by continuous SYBR Green I monitoring during amplification. *Biotechniques* **24**, 954–958, 960, 962 (1998).
34. Thisse, B. *et al.* Spatial and temporal expression of the zebrafish genome by large-scale *in situ* hybridization screening. *Methods Cell Biol.* **77**, 505–519 (2004).
35. Kolpa, H. J. *et al.* miR-21 represses Pcdcd4 during cardiac valvulogenesis. *Development* **140**, 2172–2180 (2013).
36. Tsukasaki, Y. *et al.* Giant cadherins Fat and Dachsous self-bend to organize properly spaced intercellular junctions. *Proc. Natl Acad. Sci. USA* **111**, 16011–16016 (2014).
37. Norris, R. A. *et al.* Periostin regulates atrioventricular valve maturation. *Dev. Biol.* **316**, 200–213 (2008).
38. Sauls, K. *et al.* Developmental basis for filamin-A-associated myxomatous mitral valve disease. *Cardiovasc. Res.* **96**, 109–119 (2012).
39. Xu, F., Beyazoglu, T., Hefner, E., Gurkan, U. A. & Demirci, U. Automated and adaptable quantification of cellular alignment from microscopic images for tissue engineering applications. *Tissue Eng. Part C Methods* **17**, 641–649 (2011).
40. Norris, R. A. *et al.* Identification and detection of the periostin gene in cardiac development. *Anat. Rec. A Discov. Mol. Cell. Evol. Biol.* **281A**, 1227–1233 (2004).



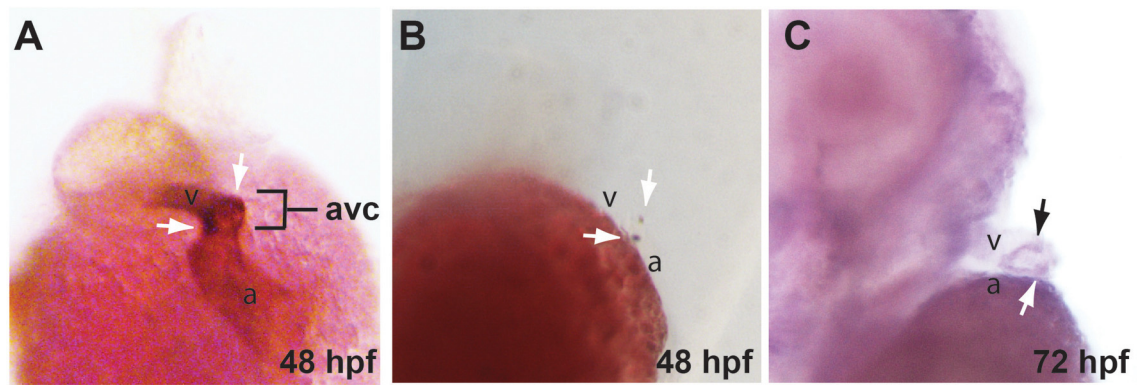
Extended Data Figure 1 | Measurement of endogenous and exogenous gene expression in *D. rerio*. **a**, Corresponding representative embryos of each morpholino knockdown on the left, with close-up of heart on the right. **b**, To assess efficiency of morpholino knockdown, 20 embryos were collected 72 h after injection, mRNA was collected, and quantitative PCR was performed with three technical replicates. We demonstrate that morpholino (MO)

knockdown of each indicated gene results in reduced mRNA expression, after normalization to beta-actin expression, compared to mock-injected controls (two-sided Student's *t*-test). *P* values are noted on graphs. **c**, Western blotting of 20 pooled embryos injected with *DCHS1* mRNA demonstrates the production of protein. Mutant mRNA refers to the compound mutant P197L/R2513H.



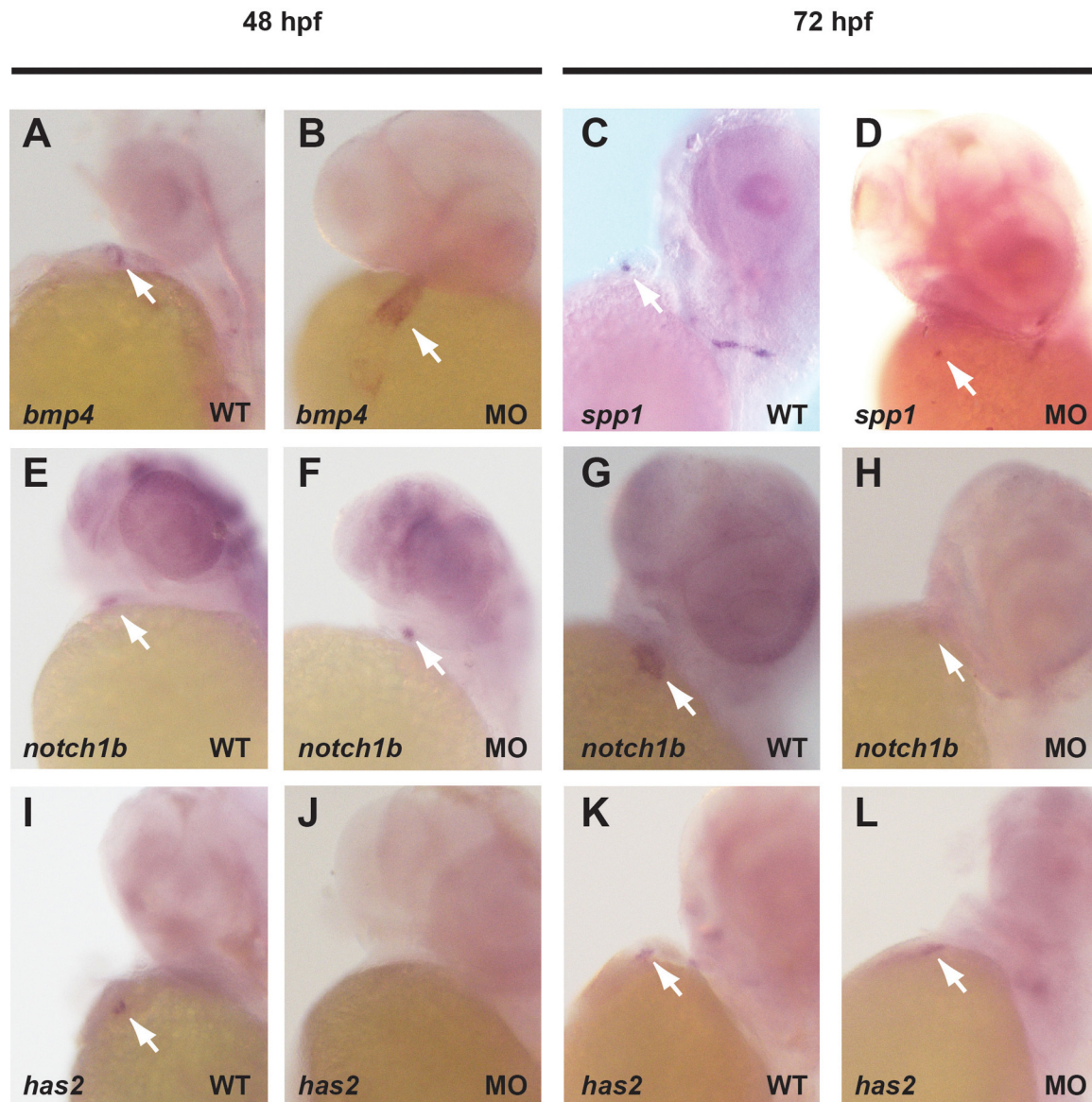
Extended Data Figure 2 | *Apbb1* is not expressed during cardiac morphogenesis. *Apbb1* RNA expression was analysed at E14.5 in sagittal sections using 2 separate antisense probes. Whereas strong cranial and neural

expression is observed for *Apbb1*, no detectable cardiac expression or valve expression (arrow) is evident.



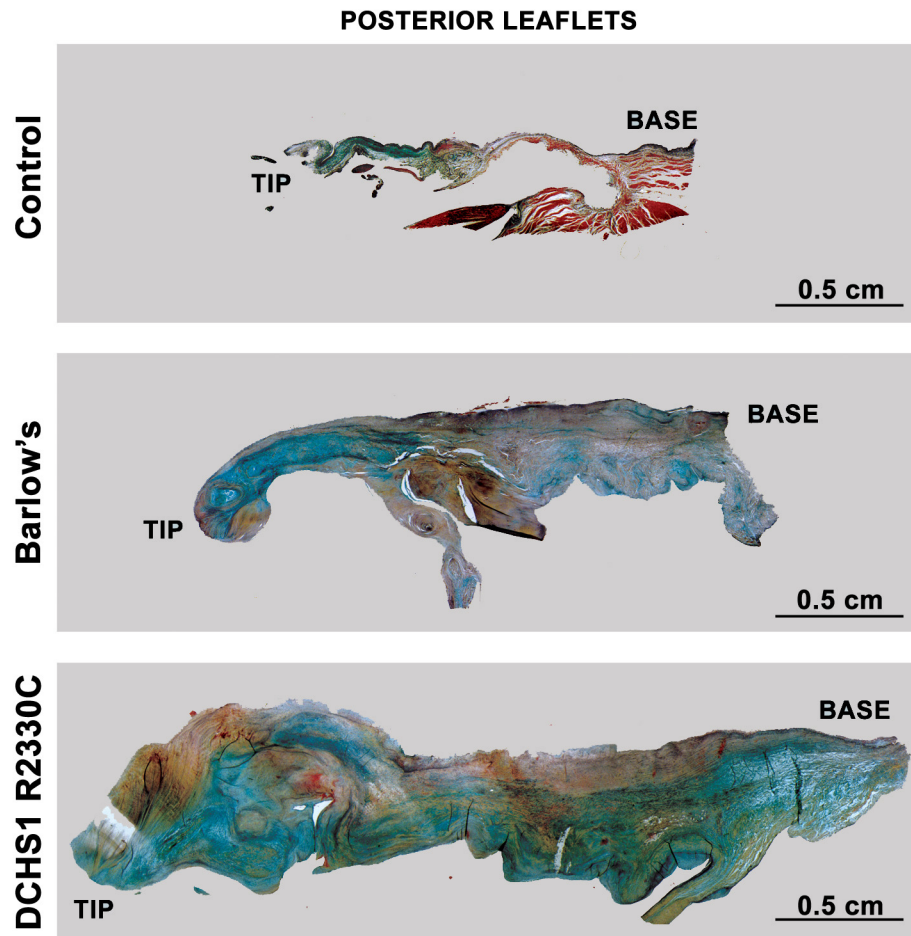
Extended Data Figure 3 | *Dachsous1b* expression at the atrioventricular junction. *In situ* hybridization reveals the presence of *dchs1b* in the atrioventricular canal (avc) at 54 hpf (**a, b**) and 72 hpf (**c**). The *dchs1b*

expression is purple while a counterstain for cardiac tissue is brown (**a**). White arrows highlight the *dchs1b* signal in the atrioventricular canal.



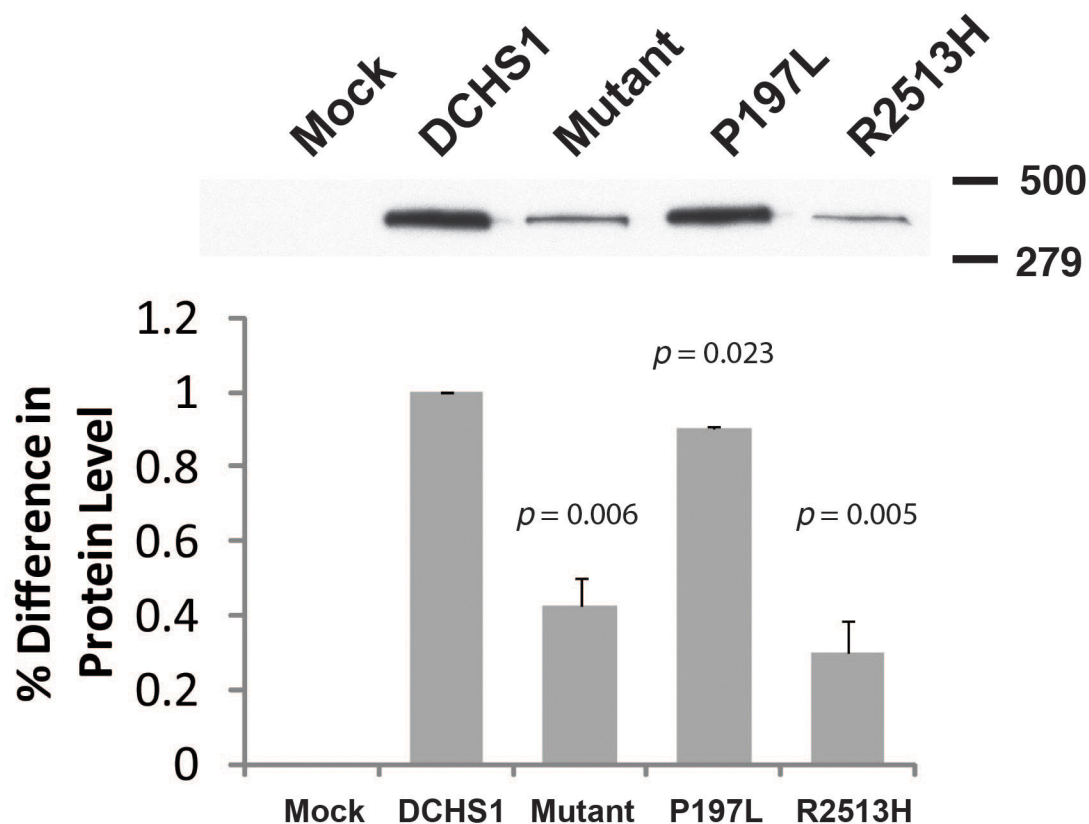
Extended Data Figure 4 | *Dachsous1b* knockdown alters atrioventricular ring markers. *In situ* hybridization at 48 hpf and 72 hpf, as indicated, was performed for known atrioventricular ring markers. In contrast to WT (a) *bmp4* expression is expanded into the ventricle at 48 hpf in *dchs1* knockdown

embryos at 48 hpf (b), *spp1* and *notch1b* expression was largely unperturbed (c–f), and *has2* expression was not detected at 48 hpf, and is faint at 72 hpf in *dchs1* knockdown, compared to identically handled and stained controls (i–l).



Extended Data Figure 5 | Histopathology mitral valves. Human posterior leaflets of control, Barlow's with MVP, and DCHS1 p.R2330C were isolated, fixed and stained with Movat's pentachrome. Leaflet thickening, elongation and myxomatous degeneration is observed in the Barlow's and DCHS1

p.R2330C leaflets compared to controls. Expansion of the proteoglycan layer (blue) and disruption of the normal stratification of matrix boundaries is observed in the Barlow's and DCHS1 p.R2330C leaflets. Blue, proteoglycan; yellow, collagen; black, elastin; red, fibrin or cardiac muscle. Scale bars, 0.5 cm.



Extended Data Figure 6 | Protein expression of uncoupled mutations. In order to determine which family 1 *DCHS1* mutation is leading to the observed decrease in protein expression, constructs were generated that harboured only the p.P197L or the p.R2513H variant. Mutant refers to the double mutant P197L/R2513H construct. Western blot analyses from transfected HEK293 cells, three independent biological replicates, demonstrate that the p.R2513H

mutation causes a significant decrease in *DCHS1* protein expression, similar to that of the construct with both variants (mutant), suggesting pathogenicity. Percent difference in protein levels is depicted. Normalization of data was accomplished by qPCR specific to the transfected constructs. *P* values from the Student's *t*-test are indicated in graphs.

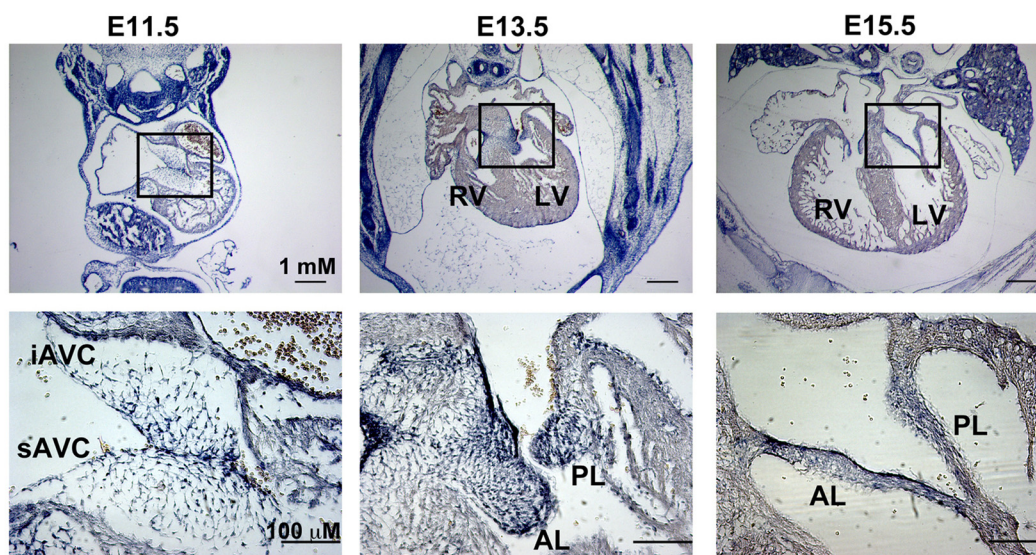
Cardiac Function (Echocardiography)

Measurement	Mode	Parameter	Units	Dchs1 +/+ (N=6)	STD	Dchs1 +/- (N=6)	STD	p-value
IVS;d	M-Mode	Depth	mm	0.8208935	0.187955	0.9096155	0.140668	0.170463
IVS;s	M-Mode	Depth	mm	1.290918667	0.243251	1.369591167	0.200513	0.393879
LVID;d	M-Mode	Depth	mm	4.282772333	0.515571	4.045031833	0.217379	0.28511
LVID;s	M-Mode	Depth	mm	2.887339833	0.33268	2.735737167	0.362674	0.544435
LVPW;d	M-Mode	Depth	mm	0.914783667	0.210393	0.961298	0.215238	0.691322
LVPW;s	M-Mode	Depth	mm	1.293790333	0.237193	1.385096	0.238861	0.252632
Calculation		Units						
EF		%		60.5315595	7.808571	61.0498945	8.235624	0.934765
FS		%		32.317489	5.734886	32.5702855	5.939884	0.955876
LV Mass		mg		148.9330448	42.74011	150.7559685	40.75343	0.891865
LV Mass (Corrected)		mg		119.1464358	34.19209	120.6047747	32.60275	0.891865
LV Vol;d		ul		83.76315083	23.58335	72.14675683	9.354702	0.261482
LV Vol;s		ul		32.456727	9.786065	28.5946545	8.983466	0.570045

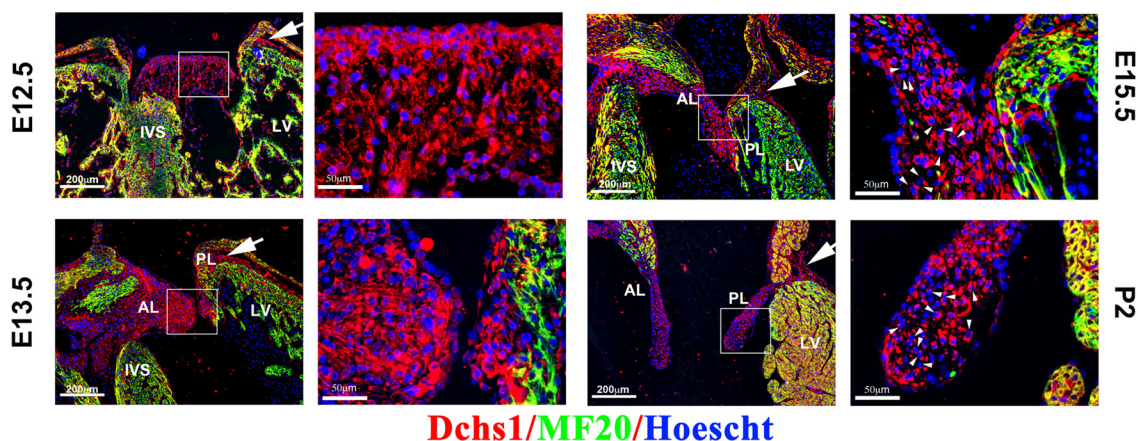
Extended Data Figure 7 | Cardiac function is not altered in *Dchs1*^{+/-} mice. M-mode analyses were performed to determine whether cardiac structure and/or function were perturbed in the *Dchs1*^{+/-} mice. No statistically significant differences were observed in either cardiac structure or calculated cardiac

function ($n = 6$ for each genotype). IVS, interventricular septum; d, diastole; s, systole; LVID, left ventricular internal dimension; LVPW, left ventricular posterior wall; EF, ejection fraction; FS, fractional shortening; LV, left ventricle.

Dchs1 mRNA Expression

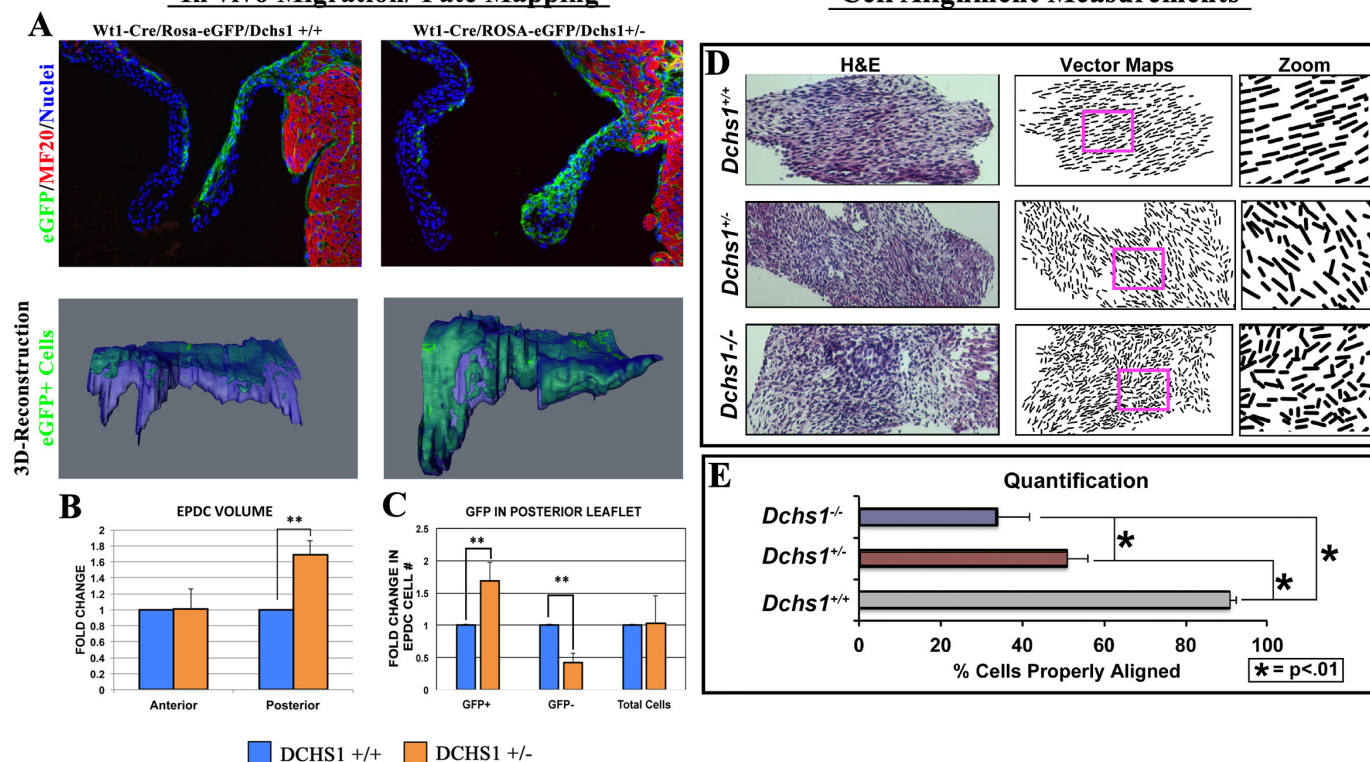


Dchs1 Protein Expression



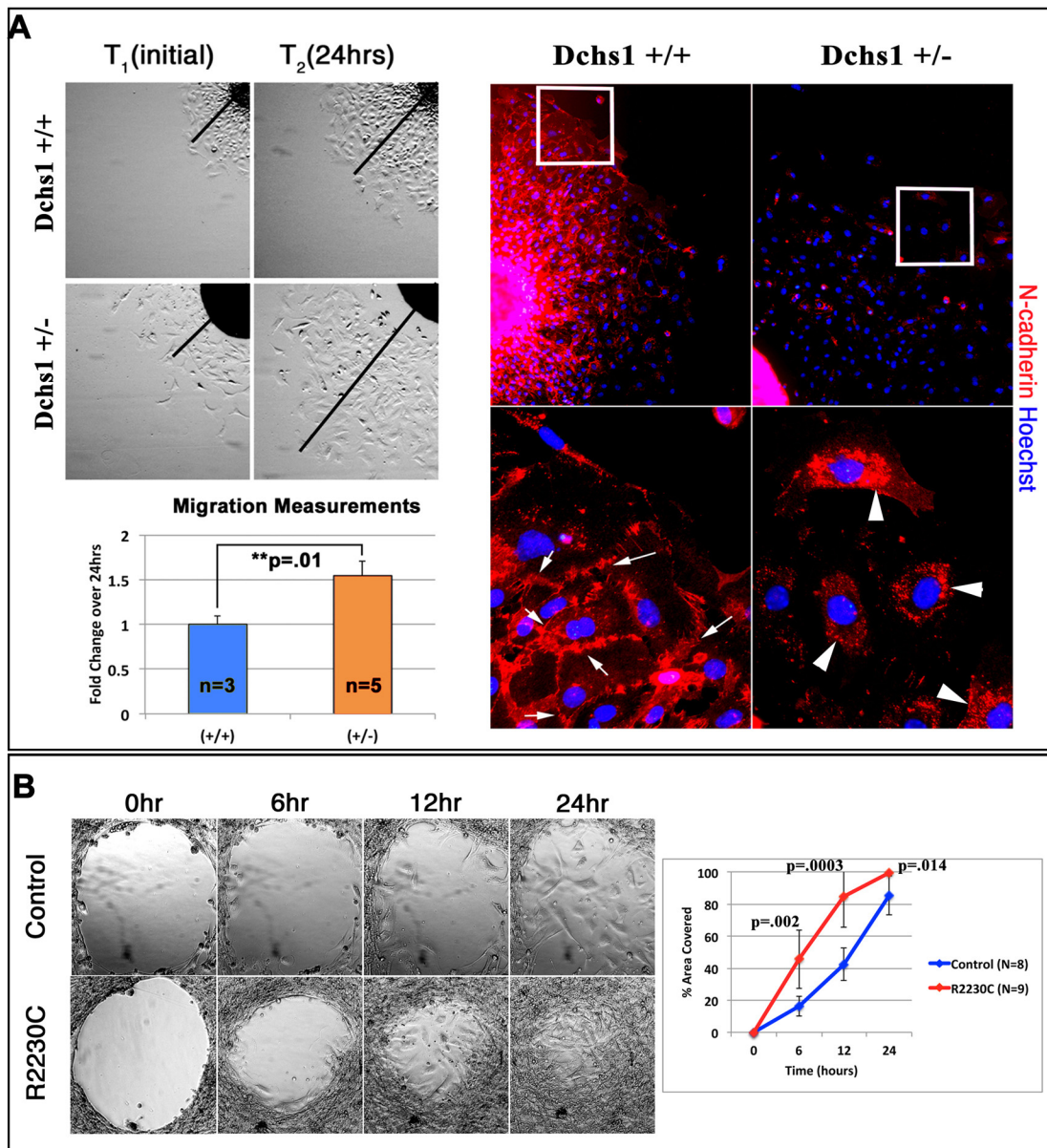
Extended Data Figure 8 | *Dchs1* expression during cardiac development. Top, RNA expression of *Dchs1* was analysed during embryonic gestation (E11.5, E13.5, and E15.5) by section *in situ* hybridization. At E11.5 *Dchs1* RNA (blue staining) expression is observed in the endocardium and mesenchyme of the superior and inferior cushions (sAVC and iAVC, respectively). A gradient pattern of expression is observed at this time point with more intense expression near the endocardium. At E13.5 and E15.5, a similar pattern

is observed in the forming anterior and posterior mitral leaflets (AL and PL, respectively). Bottom, *Dchs1* protein expression (red) is observed throughout cardiac development in the endothelial cells and interstitial cells of the developing valves. *Dchs1* shows asymmetric expression in the valvular interstitial cell bodies around E15.5 (arrowheads). *Dchs1* protein is also observed in the epicardium and atrioventricular sulcus (arrows). (Red *Dchs1*; green MF20; blue Hoescht).

In vivo Migration/ Fate Mapping**Cell Alignment Measurements**

Extended Data Figure 9 | Dchs1 deficiency causes altered valvular interstitial cell patterning *in vivo*. **a**, IHC for eGFP of postnatal day 0 (P0) lineage traced *Wt1-Cre/Rosa-eGFP/Dchs1*^{+/+} neonatal mice show epicardial-derived cells (EPDCs) migrating into the posterior leaflet as a sheet of cells directly under the endothelium of the atrialis. This normal patterning is perturbed in the *Wt1-Cre/Rosa-eGFP/Dchs1*^{+/-} mice. 3D reconstructions were used to examine all EPDCs in the posterior leaflet of both genotypes to obtain a complete fate map of these cells. **b**, **c**, Total volume of the leaflet is unchanged at this time point. However, the total volume of EPDCs as well as total EPDC cell number is significantly increased. There is a significant decrease in the number of non-EPDCs in the posterior leaflet with no overall change in total cell number. These data demonstrate that a minimum threshold of Dchs1 expression is required for normal migration of EPDCs into the

posterior leaflet, normal patterning of this cell population, and cross-talk between EPDC and non-EPDC cell types in the valve. ** $P < 0.01$. **d**, Isolated anterior mitral leaflet from fetal (E17.5) *Dchs1*^{+/+}, *Dchs1*^{+/-}, and *Dchs1*^{-/-} mice were used to quantify cellular alignment of valvular interstitial cells. Vector maps were generated from histological (haematoxylin and eosin) stains to show orientation and alignment of cells in relationship to each other. Boxes in each vector map panel are represented as zoomed images of regions within each of the valves to show cell orientation. **e**, Cell alignment and polarity were quantified as the number of cells that deviate >10 degrees from the proximal-distal (P-D) axis of the leaflet. 90% of the cells in *Dchs1*^{+/+} show proper alignment with each other and along this P-D axis. Haploinsufficiency (*Dchs1*^{+/-}) results in a 50% reduction in cell alignment, which is further reduced in *Dchs1*^{-/-} (* P values < 0.01).



Extended Data Figure 10 | Mice and MVP patients with *Dchs1* deficiency exhibit migratory defects *in vitro*. **a**, Posterior leaflets of P0 neonatal *Dchs1*^{+/+} and *Dchs1*^{+/-} mice were explanted and interstitial cells were allowed to migrate out for 24 h. *Dchs1*^{+/-} mice exhibit increased migration (black lines drawn from explants) coincident with loss of cell–cell contacts and N-cadherin expression at focal adhesions. Whereas N-cadherin expression (red) is found at the membrane at points of cell–cell contact in *Dchs1*^{+/+}

valvular interstitial cells (arrows), this membrane expression is lost in the *Dchs1*^{+/-} cells and is prominently expressed in the cytoplasm (arrows). Nuclei, blue. **b**, Migration assays using control and MVP patient (p.R2230C) valvular interstitial cells exhibit a similar affect as observed in the mouse cells whereby the p.R2230C cells exhibit an increase in migration. *P* values are indicated in graphs.

PIK3CA^{H1047R} induces multipotency and multi-lineage mammary tumours

Shany Koren¹, Linsey Reavie¹, Joana Pinto Couto¹, Duvini De Silva¹, Michael B. Stadler^{1,2}, Tim Roloff¹, Adrian Britschgi¹, Tobias Eichlisberger¹, Hubertus Kohler¹, Olulanu Aina³, Robert D. Cardiff³ & Mohamed Bentes-Alj¹

The adult mouse mammary epithelium contains self-sustained cell lineages that form the inner luminal and outer basal cell layers, with stem and progenitor cells contributing to its proliferative and regenerative potential^{1–4}. A key issue in breast cancer biology is the effect of genomic lesions in specific mammary cell lineages on tumour heterogeneity and progression. The impact of transforming events on fate conversion in cancer cells of origin and thus their contribution to tumour heterogeneity remains largely elusive. Using *in situ* genetic lineage tracing and limiting dilution transplantation, we have unravelled the potential of PIK3CA^{H1047R}, one of the most frequent mutations occurring in human breast cancer⁵, to induce multipotency during tumorigenesis in the mammary gland. Here we show that expression of PIK3CA^{H1047R} in lineage-committed basal Lgr5-positive and luminal keratin-8-positive cells of the adult mouse mammary gland evokes cell dedifferentiation into a multipotent stem-like state, suggesting this to be a mechanism involved in the formation of heterogeneous, multi-lineage mammary tumours. Moreover, we show that the tumour cell of origin influences the frequency of malignant mammary tumours. Our results define a key effect of PIK3CA^{H1047R} on mammary cell fate in the pre-neoplastic mammary gland and show that the cell of origin of PIK3CA^{H1047R} tumours dictates their malignancy, thus revealing a mechanism underlying tumour heterogeneity and aggressiveness.

The mammary gland epithelium is composed of two major cell lineages: the luminal layer contains cells expressing keratin 8/18 (K8/18) and the basal layer with cells expressing K5/14 and/or smooth muscle actin (SMA) and/or p63 (ref. 6).

Multipotent cells that generate both the luminal and basal lineages are found in the mouse embryonic mammary gland^{1,7} but their existence in the adult gland is still under debate. Studies using serial transplantation into cleared mammary fat pad supposed the existence in the adult mouse mammary gland of multipotent stem cells with myo-epithelial features^{8–11}. Arguably, these assays reflected the regenerative potential of the transplanted cells rather than their properties *in situ*^{1–3}. Lineage-tracing studies, which permit targeted expression of a fluorescent reporter in a given cell and its progeny, showed that tissue homeostasis is maintained by unipotent luminal K8/18-positive and basal K5/14/Lgr5-positive stem cells after birth. Lineage tracing of K8/18, K5/14 and Lgr5 progeny found no evidence for the presence of multipotent stem cells in the adult mammary gland¹ but did not exclude the possibility that rare cells not targeted by these reporters, or only at a very low frequency, have multipotent potential. While tracing of the progeny of axin-2-positive cells showed the presence of multipotent stem cells during puberty and pregnancy, this and other studies revealed that the basal and luminal lineages are self-sustained in the adult virgin gland^{2–4}. By contrast, recent three-dimensional whole-mount imaging¹² and the identification of the Procr-positive subset¹³ argue for the presence of multipotent stem cells in the adult virgin mouse mammary gland, thus reopening the debate.

The phosphatidylinositol 3-kinase (PI3K) pathway is activated in ~70% of breast cancers. Several mechanisms may account for the activation of this pathway in cancer, including amplification and/or activating mutations of the PIK3CA gene that encodes the p110 α catalytic subunit of PI3K found in 20–40% of breast cancers^{5,14}. The most recurrent mutation, H1047R, leads to constitutive PI3K signaling and heterogeneous mammary tumours^{15–17}. Despite frequent alterations of the PI3K pathway in breast cancer, its impact on lineage organization during tumorigenesis and the importance of the cell of origin for heterogeneity and aggressiveness of PI3K-driven tumours has remained unclear.

To address the effects of mutant PIK3CA^{H1047R} on basal- or luminal-lineage-restricted cells, we performed lineage tracing in adult Lgr5-CreER^{T2}/Tomato-reporter and K8-CreER^{T2}/Tomato-reporter mice with or without PIK3CA^{H1047R} (Extended Data Fig. 1a)¹⁶. Moreover, we used Lgr5- and K8-CreER^{T2}/PIK3CA^{H1047R} or PIK3CA wild-type (PIK3CA^{WT}) animals^{16,18} for tracing of the green fluorescent protein (GFP) reporter, and Lgr5-CreER^{T2} and K8-CreER^{T2} mice as controls (Extended Data Fig. 1b). As previously reported^{1,19}, we found Lgr5 activity only in a subset of basal cells in the nipple area of the mammary gland (Extended Data Fig. 2a–e). We further assessed the effects of PIK3CA^{H1047R} expression on the distribution of mammary subpopulations by fluorescence-activated cell sorting (FACS) on isolated mammary epithelial cells labelled with CD24 and Sca1, which were shown to enrich for luminal (CD24^{Hi}Sca1[–], CD24^{Hi}Sca1⁺) and basal (CD24^{Lo}Sca1[–]) cells^{10,20} (Extended Data Fig. 3). Four days lineage tracing confirmed Tomato labelling of basal cells in Lgr5-CreER^{T2}/Tomato and Lgr5-CreER^{T2}/PIK3CA^{H1047R}/Tomato and of luminal cells in K8-CreER^{T2}/Tomato and K8-CreER^{T2}/PIK3CA^{H1047R}/Tomato animals (Extended Data Figs 2b, f and 4a, b). We further performed 4-, 8- and 13-week lineage tracing. The distribution of Tomato-labelled subsets in Lgr5-CreER^{T2}/Tomato glands did not change with time where the progeny of Lgr5-positive cells was mostly of basal origin (Fig. 1a, b and Extended Data Fig. 2g, h). In K8-CreER^{T2}/Tomato mice, labelling was restricted to the luminal subset, marking mostly mature CD24^{Hi}Sca1⁺ in 4 weeks and mostly CD24^{Hi}Sca1[–] luminal progenitors in 13 weeks tracing, indicating the targeting of a long-term unipotent luminal subset (Fig. 1a, c and Extended Data Fig. 4c, d). By contrast, expression of PIK3CA^{H1047R} in Lgr5-CreER^{T2}/PIK3CA^{H1047R}/Tomato and K8-CreER^{T2}/PIK3CA^{H1047R}/Tomato mice resulted in labelling of both the luminal and the basal compartments (Fig. 1b, c and Extended Data Figs 2g, h, 4c, d). PIK3CA^{H1047R}-evoked multi-lineage labelling was not observed at 4–7 days after tamoxifen induction (Extended Data Figs 2f, i and 4b). Since PIK3CA^{H1047R/WT} targeting vectors contain an internal ribosome entry site (IRES)–GFP construct, we also used GFP as a readout of transgene expression. At 4 days after tamoxifen induction, K8-CreER^{T2}/PIK3CA^{WT} and K8-CreER^{T2}/PIK3CA^{H1047R} glands expressed similar levels of GFP, indicating similar Cre recombination efficiency. In both models, 4- and 8–11-week lineage tracing revealed an increase in GFP-labelled

¹Friedrich Miescher Institute for Biomedical Research (FMI), 4058 Basel, Switzerland. ²Swiss Institute of Bioinformatics, 4058 Basel, Switzerland. ³Department of Pathology, Center for Comparative Medicine, University of California Davis, Davis, California 95616, USA.

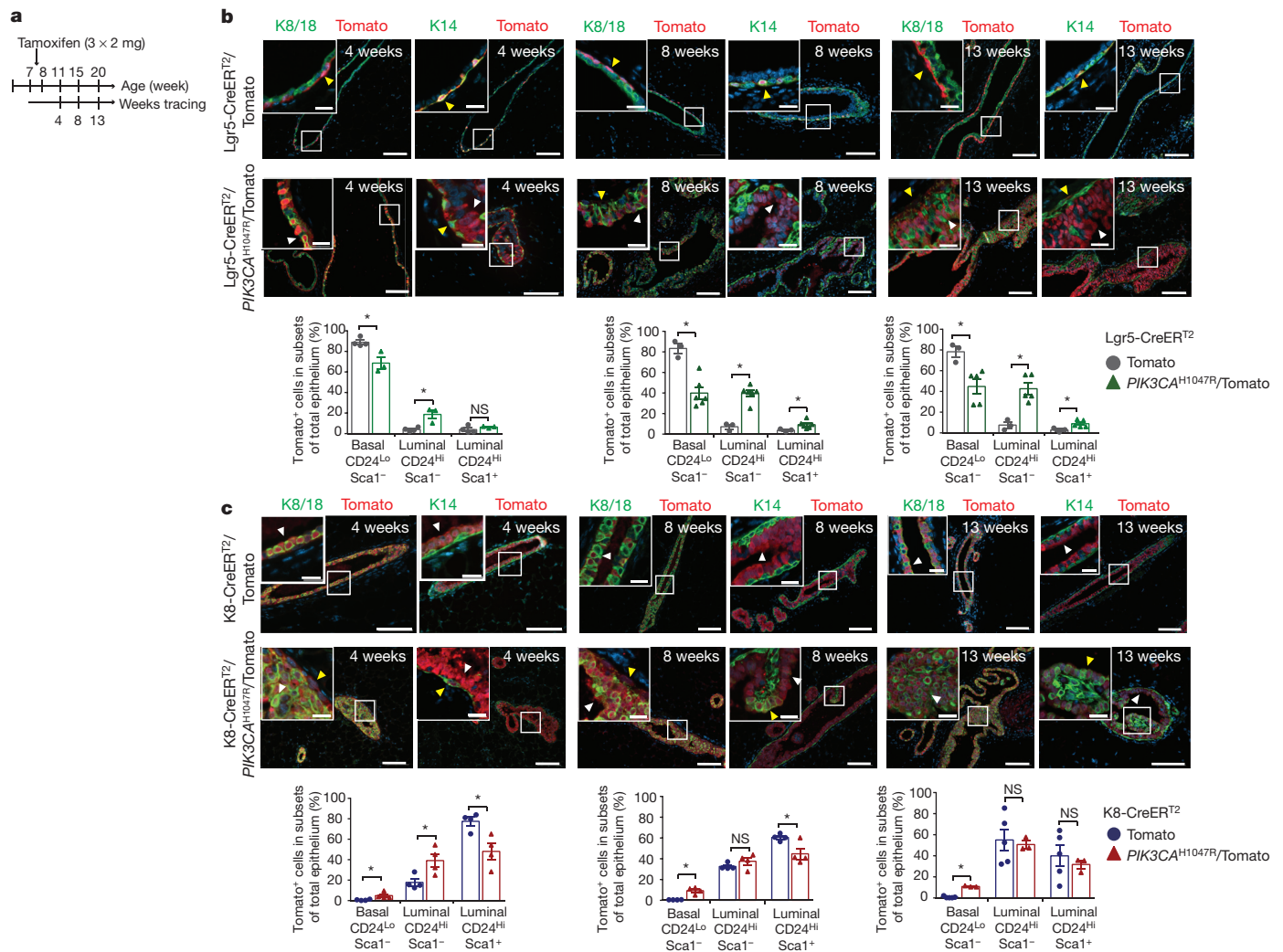


Figure 1 | Mutant *PIK3CA* induces mammary cell plasticity. **a**, Timeline for lineage-tracing studies. **b, c**, Representative images of 4-, 8- and 13-week tracing and FACS quantification of Tomato-positive epithelial basal (CD24^{Lo}Sca1⁻) and luminal (CD24^{Hi}Sca1⁺) subsets from Lgr5-CreER^{T2}/Tomato (b, immunofluorescence: left *n* = 3, middle *n* = 11, right *n* = 6 mice; FACS: left *n* = 4 technical replicates (each 1–3 pooled mice), middle *n* = 3 technical replicates (each 1–2 pooled mice), right *n* = 3 technical replicates (each 1 mouse)), Lgr5-CreER^{T2}/PIK3CA^{H1047R}/Tomato (b, immunofluorescence: left *n* = 3, middle *n* = 10, right *n* = 4 mice; FACS: left *n* = 3 technical replicates (each 1–2 pooled mice), middle *n* = 6 technical replicates (each 1 mouse), right *n* = 5 technical replicates (each 1–2 pooled mice)), K8-CreER^{T2}/Tomato

(c, immunofluorescence: left *n* = 4, middle *n* = 8, right *n* = 5 mice; FACS: left *n* = 4 technical replicates (each 1–3 pooled mice), middle *n* = 4 technical replicates (each 1–2 pooled mice), right *n* = 5 technical replicates (each 1 mouse)) and K8-CreER^{T2}/PIK3CA^{H1047R}/Tomato animals (c, immunofluorescence: left *n* = 4, middle *n* = 4, right *n* = 3 mice; FACS: left *n* = 4 technical replicates (each 1–2 pooled mice), middle *n* = 4, right *n* = 3 technical replicates (each 1 mouse)). White arrowheads indicate luminal and yellow arrowheads indicate basal Tomato-labelled cells. Scale bars, 100 μm, 20 μm (magnifications). Bar graphs show means ± standard error of the mean (s.e.m.); two-sided unpaired Student's *t*-test; **P* < 0.05; NS, not significant.

basal and luminal subsets in PIK3CA^{H1047R} compared with PIK3CA^{WT} animals that showed lineage-restricted GFP labelling, consistent with higher PI3K pathway activation (Extended Data Figs 5 and 6). These results suggest that expression of PIK3CA^{H1047R} in basal- or luminal-restricted mammary cells triggers lineage plasticity and cell expansion.

Next, we assessed the effects of PIK3CA^{H1047R} on global gene expression in the pre-neoplastic gland. Microarray and quantitative polymerase chain reaction with reverse transcription (qRT-PCR) analyses were performed on mammary epithelial cell subpopulations from Lgr5-CreER^{T2} control versus Lgr5-CreER^{T2}/PIK3CA^{H1047R} and K8-CreER^{T2} control versus K8-CreER^{T2}/PIK3CA^{H1047R} animals. Genes expressed differentially between subpopulations of control and mutant mice were compared with subpopulation signatures that have been previously described^{21,22}. We found enrichment of luminal progenitor signature genes in the basal Lgr5-CreER^{T2}/PIK3CA^{H1047R} subset and in the newly formed basal K8-CreER^{T2}/PIK3CA^{H1047R} subset. Enrichment of myoepithelial signature genes was found in the newly

formed luminal Lgr5-CreER^{T2}/PIK3CA^{H1047R} and in the K8-CreER^{T2}/PIK3CA^{H1047R} luminal subsets (Fig. 2a, b and Extended Data Fig. 7). Mammary cells co-expressing basal and luminal markers in neoplastic areas confirmed that PIK3CA^{H1047R} induces cell plasticity (Fig. 2c).

We addressed how lineage plasticity is evoked by PIK3CA^{H1047R} in functional assays. Limiting dilution transplantation revealed an increase in the mammary-repopulating capacity of basal cells from Lgr5-CreER^{T2}/PIK3CA^{H1047R} and K8-CreER^{T2}/PIK3CA^{H1047R} mice compared with the respective controls (Fig. 3a). The GFP-negative basal CD24^{Lo}Sca1⁻ population from K8-CreER^{T2} mice comprises the bulk of basal cells (myoepithelial and mammary-repopulating cells)^{10,20}, explaining the lack of outgrowths at the number of cells transplanted. PIK3CA^{H1047R}-expressing luminal cells also had repopulating capacity (Extended Data Fig. 8a, b). All outgrowths expressed luminal and basal markers (Fig. 3b and Extended Data Fig. 8c). In colony formation assays, PIK3CA^{H1047R} increased the percentage of double-positive (K14/K8/18) colonies derived from the newly formed

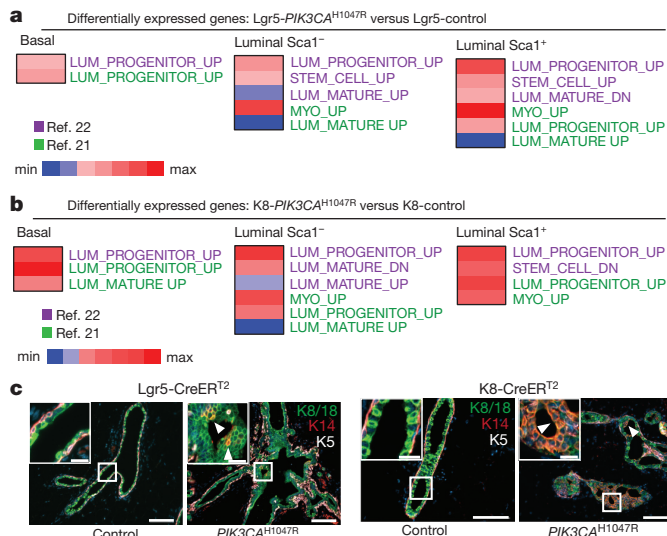


Figure 2 | Activation of *PIK3CA*^{H1047R} leads to expression of basal- and luminal-lineage genes. **a**, **b**, Plots indicating enrichment of gene expression from FACS-sorted Lgr5-CreER^{T2}/*PIK3CA*^{H1047R} versus Lgr5-CreER^{T2} control (**a**) and K8-CreER^{T2}/*PIK3CA*^{H1047R} versus K8-CreER^{T2} control subsets (**b**) in signatures of mammary subpopulations from refs 21, 22. Microarray was performed 4 weeks after tamoxifen induction on basal CD24^{Lo}Sca1⁻ and luminal CD24^{Hi}Sca1⁺ subsets of pooled mammary glands of 2–3 oestrus-synchronized animals from three independent sortings. **c**, Representative images of immunostaining for basal (K14, red; K5, white) and luminal (K8/18, green) markers on mammary glands 4 weeks after tamoxifen treatment ($n = 3$ mice). White arrowheads indicate double-positive cells. Scale bars, 100 μ m, 20 μ m (magnifications).

luminal (Lgr5-CreER^{T2}/*PIK3CA*^{H1047R}) and basal cells (K8-CreER^{T2}/*PIK3CA*^{H1047R}). *PIK3CA*^{H1047R}-expressing subsets overcame lineage restriction, giving rise to both lineages, albeit at low frequencies. Moreover, basal cells from Lgr5- and K8-CreER^{T2}/*PIK3CA*^{H1047R} animals showed increased colony formation capacity (Extended Data Fig. 8d–f). In mammosphere cultures, *PIK3CA*^{H1047R} increased the sphere-forming capacity of luminal cells. While control cells formed spheres with a hollow lumen after passaging, *PIK3CA*^{H1047R}-expressing cells formed filled spheres, indicating the accumulation of less differentiated cells (Extended Data Fig. 8g, h). Altogether, these data suggest that *PIK3CA*^{H1047R} evokes cell dedifferentiation to a multipotent stem-like state, from which cells further differentiate into both cell lineages.

Expression of *PIK3CA*^{H1047R} in Lgr5- and K8-positive cells induced mammary tumours on average after 108 and 78 days, respectively. Control and *PIK3CA*^{WT} animals developed no tumours (Fig. 4a). FACS analysis of GFP-positive tumour cells revealed a similar distribution of cancer subpopulations, with an accumulation of the CD24^{Hi}Sca1⁻ subset. *PIK3CA*^{H1047R}-evoked mammary tumours expressed basal and luminal markers (Extended Data Fig. 9a–d). Additionally, cells double positive for basal and luminal markers were found (Fig. 4b). These results suggest that *PIK3CA*^{H1047R}-evoked cell plasticity results in multi-lineage tumours and that expression of basal and luminal markers is not an indicator of the origin of mammary cancers. It has been proposed that basal-like mammary tumours may originate from luminal cells^{23,24}. Therefore, any inference of the cell of origin from the differentiation state of the tumour can be misleading.

Histological analysis showed heterogeneous phenotypes and differences in malignancy between both models. Lgr5-CreER^{T2}/*PIK3CA*^{H1047R} mice formed unique benign multi-nodular rosette-type adenomyoepitheliomas, aggressive adenosquamous carcinomas with pilosebaceous differentiation and carcinosarcomas. K8-CreER^{T2}/*PIK3CA*^{H1047R} mice mainly developed aggressive adenosquamous carcinoma, carcinosarcomas that infiltrated the surrounding tissue, adenocarcinomas and benign adenomyoepitheliomas. These results

a Lgr5-CreER^{T2} outgrowths

	Cell no.	Control	<i>PIK3CA</i> ^{H1047R}
Basal	50	5/17 (29%)	10/19 (52%)
CD24 ^{Lo}	100	4/15 (26%)	9/15 (60%)
Sca1 ⁻	200	5/13 (38%)	9/13 (69%)
MRU frequency (95% CI)		1 in 289 (1/169–1/496)	1 in 111 (1/74–1/166)

$P = 0.003$

K8-CreER^{T2} outgrowths

	Cell no.	Control	<i>PIK3CA</i> ^{H1047R}
Basal	100	0/6 (0%)	4/10 (40%)
CD24 ^{Lo}	200	0/8 (0%)	7/11 (63%)
Sca1 ⁻	400	0/5 (0%)	2/4 (50%)
MRU frequency (95% CI)		N/A	1 in 253 (1/142–1/452)

$P = 4 \times 10^{-6}$

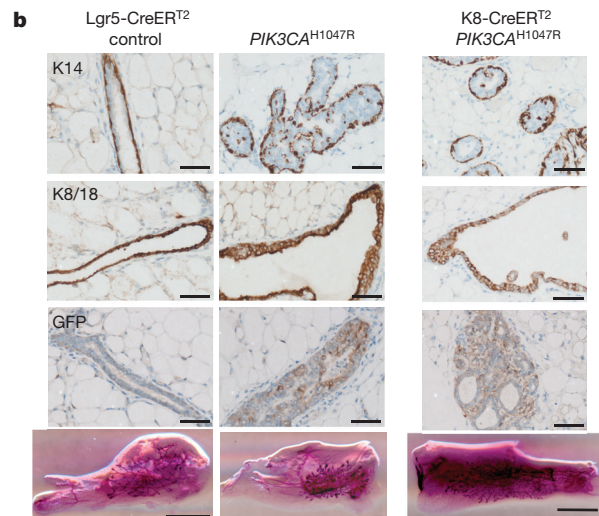


Figure 3 | Expression of *PIK3CA*^{H1047R} evokes multipotent stem-like cells. **a**, Number of outgrowths in cleared-fat-pad transplantation of GFP-positive (Lgr5-positive) control and *PIK3CA*^{H1047R}-expressing basal subsets (top). Outgrowths of GFP-negative control and GFP-positive *PIK3CA*^{H1047R}-expressing basal subsets (bottom). The GFP-negative control basal subset comprises mammary repopulating cells and mostly myoepithelial cells. CI, confidence interval; MRU, mammary repopulating unit; N/A, not applicable. **b**, Representative immuno-stained sections ($n = 3$ mice) and carmine-stained whole mounts of outgrowths from cleared-fat-pad transplantation (100 cells condition). Scale bars, 50 μ m (top three rows), 500 μ m (bottom row). **a**, **b**, Pooled data from three independent experiments.

show that *PIK3CA*^{H1047R} mostly evokes benign tumours (74%) with high intratumour heterogeneity when expressed in Lgr5-positive cells, in contrast to the mostly aggressive mammary tumours (62%) with a distinctive infiltrative densely fibrotic phenotype seen when the cell of origin is K8-positive (Fig. 4c and Extended Data Fig. 9e, f).

Microarray analysis, principle component analysis and hierarchical clustering of tumours revealed a single cluster of Lgr5-CreER^{T2}/*PIK3CA*^{H1047R} and three clusters of K8-CreER^{T2}/*PIK3CA*^{H1047R} tumours. We found no correlation between tumour phenotype and clustering, probably owing to intratumour heterogeneity. We compared tumour expression profiles to different human breast cancer subtypes⁵. The majority (7/10) of the K8-CreER^{T2}/*PIK3CA*^{H1047R} tumours correlated best with the malignant basal-like, HER2-enriched and luminal-B profiles, whereas 3/10 clustered with benign luminal-A and normal-like breast cancers. Lgr5-CreER^{T2}/*PIK3CA*^{H1047R} tumours resembled benign (2/10) but also malignant subtypes (3/10). However, 5/10 showed no similarity to a single subtype but were equidistant from all, representing the high intratumour heterogeneity in this model (Fig. 4d and Extended Data Fig. 10). The fact that

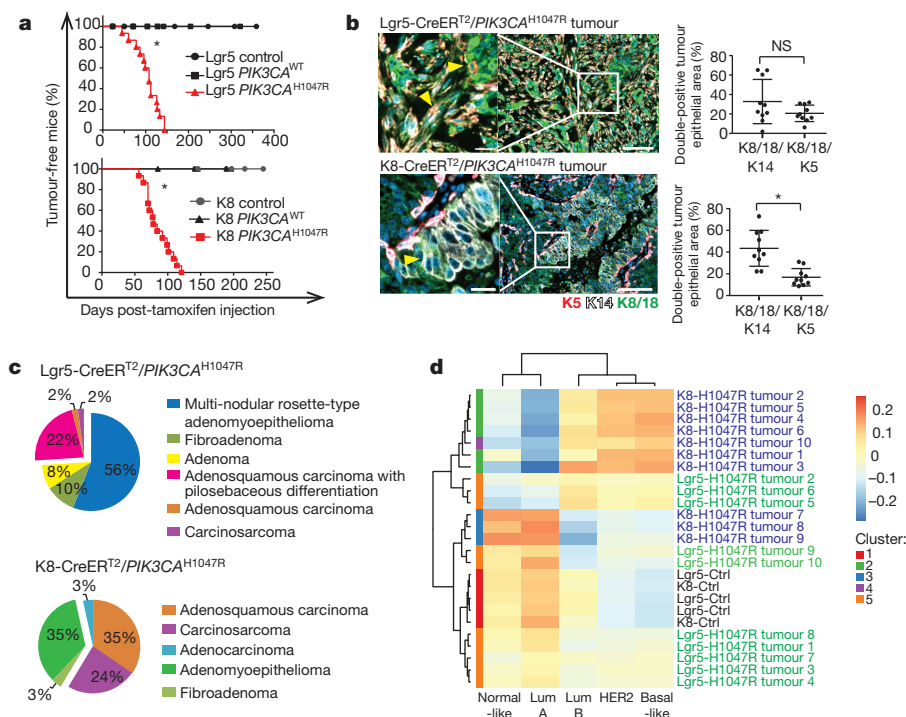


Figure 4 | The frequency of malignant tumour lesions is dictated by the cell of origin. **a**, Kaplan-Meier plots depicting tumour onset in Lgr5- $\text{CreER}^{T2}/\text{PIK3CA}^{\text{H1047R}}$ (average 108 days; $n = 15$) and K8- $\text{CreER}^{T2}/\text{PIK3CA}^{\text{H1047R}}$ (average 78 days; $n = 15$) mice compared with $\text{PIK3CA}^{\text{WT}}$ and control animals after tamoxifen injection (top, $n = 14$; bottom, $n = 9$ and $n = 12$, respectively). * $P < 0.0001$, log-rank test. **b**, Representative images of immunostaining and quantification of double-positive epithelial area in Lgr5- $\text{CreER}^{T2}/\text{PIK3CA}^{\text{H1047R}}$ and K8- $\text{CreER}^{T2}/\text{PIK3CA}^{\text{H1047R}}$ tumours ($n = 10$). Yellow arrowheads indicate double-positive cells. Scale bars, 100 μm (right), 20 μm (left, magnification). Graphs show means \pm standard deviation (s.d.); two-sided unpaired Student's t -test; * $P = 0.0002$; NS, not significant. **c**, Pie chart showing phenotypes of tumours evoked in Lgr5- $\text{CreER}^{T2}/\text{PIK3CA}^{\text{H1047R}}$ ($n = 50$) and K8- $\text{CreER}^{T2}/\text{PIK3CA}^{\text{H1047R}}$ ($n = 29$) animals. **d**, Clustered heat map showing correlation coefficients between human breast cancer profiles and ten K8- and Lgr5- $\text{CreER}^{T2}/\text{PIK3CA}^{\text{H1047R}}$ mouse mammary tumours. Ctrl, control; Lum, luminal.

K8- $\text{CreER}^{T2}/\text{PIK3CA}^{\text{H1047R}}$ tumours but not Lgr5- $\text{CreER}^{T2}/\text{PIK3CA}^{\text{H1047R}}$ tumours clustered mostly with malignant breast cancers that have a poor prognosis is consistent with the histopathological results and suggests that, in the presence of the same initiating oncogenic mutation, the cell of origin dictates the frequency of aggressive tumours.

GFP driven by the Lgr5 promoter was found in a rare luminal subset of the control gland (Extended Data Fig. 5b, c), as shown previously¹. It is unlikely that such rare cells expand due to $\text{PIK3CA}^{\text{H1047R}}$ and form the luminal GFP-positive population. We observed multi-lineage labelling upon $\text{PIK3CA}^{\text{H1047R}}$ expression in a basal and a luminal cell-driven model. An alternative possibility that we cannot firmly exclude is that Lgr5- and K8-positive populations contain rare bipotent subsets that are quiescent or not efficiently labelled in physiological conditions and are, therefore, not detected by lineage tracing but may expand upon $\text{PIK3CA}^{\text{H1047R}}$ expression.

We show that expression of $\text{PIK3CA}^{\text{H1047R}}$ dedifferentiates lineage-restricted epithelial cells into a multipotent stem-like state from which cells further differentiate, revealing a mechanism by which heterogeneous mixed-lineage tumours may develop. Furthermore, we show that the tumour cell of origin influences the frequency of malignant mammary tumours. The fundamental questions of which mammary cells are susceptible to which combination of oncogenes and how this impinges on tumour progression and aggressiveness warrant further investigation. Understanding these dynamic relationships is paramount for understanding tumour heterogeneity and for identifying prognostic and predictive biomarkers.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 30 July 2014; accepted 16 June 2015.

Published online 12 August; corrected online 2 September 2015 (see full-text HTML version for details).

- Van Keymeulen, A. *et al.* Distinct stem cells contribute to mammary gland development and maintenance. *Nature* **479**, 189–193 (2011).
- van Amerongen, R., Bowman, A. N. & Nusse, R. Developmental stage and time dictate the fate of Wnt/ β -catenin-responsive stem cells in the mammary gland. *Cell Stem Cell* **11**, 387–400 (2012).

- de Visser, K. E. *et al.* Developmental stage-specific contribution of LGR5⁺ cells to basal and luminal epithelial lineages in the postnatal mammary gland. *J. Pathol.* **228**, 300–309 (2012).
- Tao, L., van Bragt, M. P., Laudadio, E. & Li, Z. Lineage tracing of mammary epithelial cells using cell-type-specific Cre-expressing adenoviruses. *Stem Cell Rep.* **2**, 770–779 (2014).
- The Cancer Genome Atlas Network. Comprehensive molecular portraits of human breast tumours. *Nature* **490**, 61–70 (2012).
- Hennighausen, L. & Robinson, G. W. Signaling pathways in mammary gland development. *Dev. Cell* **1**, 467–475 (2001).
- Spike, B. T. *et al.* A mammary stem cell population identified and characterized in late embryogenesis reveals similarities to human breast cancer. *Cell Stem Cell* **10**, 183–197 (2012).
- Shackleton, M. *et al.* Generation of a functional mammary gland from a single stem cell. *Nature* **439**, 84–88 (2006).
- Stingl, J. *et al.* Purification and unique properties of mammary epithelial stem cells. *Nature* **439**, 993–997 (2006).
- Sleeman, K. E. *et al.* Dissociation of estrogen receptor expression and *in vivo* stem cell activity in the mammary gland. *J. Cell Biol.* **176**, 19–26 (2007).
- Prater, M. D. *et al.* Mammary stem cells have myoepithelial cell properties. *Nature Cell Biol.* **16**, 942–950 (2014).
- Rios, A. C., Fu, N. Y., Lindeman, G. J. & Visvader, J. E. *In situ* identification of bipotent stem cells in the mammary gland. *Nature* **506**, 322–327 (2014).
- Wang, D. *et al.* Identification of multipotent mammary stem cells by protein C receptor expression. *Nature* **517**, 81–84 (2015).
- Samuels, Y. *et al.* High frequency of mutations of the PIK3CA gene in human cancers. *Science* **304**, 554 (2004).
- Koren, S. & Bentires-Alj, M. Mouse models of PIK3CA mutations: one mutation initiates heterogeneous mammary tumors. *FEBS J.* **280**, 2758–2765 (2013).
- Meyer, D. S. *et al.* Luminal expression of PIK3CA mutant H1047R in the mammary gland induces heterogeneous tumors. *Cancer Res.* **71**, 4344–4351 (2011).
- Liu, P. *et al.* Oncogenic PIK3CA -driven mammary tumors frequently recur via PI3K pathway-dependent and PI3K pathway-independent mechanisms. *Nature Med.* **17**, 1116–1120 (2011).
- Meyer, D. S. *et al.* Expression of PIK3CA mutant E545K in the mammary gland induces heterogeneous tumors but is less potent than mutant H1047R. *Oncogenesis* **2**, e74 (2013).
- Plaks, V. *et al.* Lgr5-expressing cells are sufficient and necessary for postnatal mammary gland organogenesis. *Cell Rep.* **3**, 70–78 (2013).
- Sleeman, K. E., Kendrick, H., Ashworth, A., Isacke, C. M. & Smalley, M. J. CD24 staining of mouse mammary gland cells defines luminal epithelial, myoepithelial/basal and non-epithelial cells. *Breast Cancer Res.* **8**, R7 (2006).
- Meier-Abt, F. *et al.* Parity induces differentiation and reduces Wnt/Notch signaling ratio and proliferation potential of basal stem/progenitor cells isolated from mouse mammary epithelium. *Breast Cancer Res.* **15**, R36 (2013).
- Lim, E. *et al.* Transcriptome analyses of mouse and human mammary cell subpopulations reveal multiple conserved genes and pathways. *Breast Cancer Res.* **12**, R21 (2010).

23. Lim, E. *et al.* Aberrant luminal progenitors as the candidate target population for basal tumor development in *BRCA1* mutation carriers. *Nature Med.* **15**, 907–913 (2009).
24. Molyneux, G. *et al.* *BRCA1* basal-like breast cancers originate from luminal epithelial progenitors and not from basal stem cells. *Cell Stem Cell* **7**, 403–417 (2010).

Acknowledgements The authors thank R. Thierry for help with image processing and quantification, S. Bichet and A. Bogucki for assistance with immunohistochemical staining, S. Thiry for assistance with the microarray analysis, L. Gelman, S. Bourke and M. Kirschmann for help with microscopy, C. Blanpain, B. Roska, B. Kinzel, J. Tchorz and A. Isken for providing mouse lines, and members of the Bentires-Alj group for their feedback. O.A. and R.D.C. were supported by National Cancer Institute grant U01 CA141582. Research in the laboratory of M.B.-A. is supported by the Novartis Research Foundation, the European Research Council (ERC starting grant 243211-PTPsBDC), the Swiss Cancer League, the Swiss National Foundation, and the Krebsliga Beider Basel.

Author Contributions S.K. and M.B.-A. designed experiments, analysed the data and wrote the manuscript. L.R. contributed greatly to experimental design and data analysis. S.K. performed most of the experiments. L.R. and D.D.S. performed mammosphere cultures. L.R., D.D.S., J.P.C. and S.K. performed limiting dilution transplantations. M.B.S. and T.R. performed microarray data analysis. L.R. and J.P.C. quantified tumour immunohistochemistry. J.P.C. and A.B. isolated tumour RNA. A.B. performed immunoblotting. D.D.S. and T.E. provided technical assistance for several experiments. H.K. provided technical assistance for FACS experiments. O.A. and R.D.C. analysed histological tumour samples. All the authors discussed the data and participated in the preparation of the manuscript.

Author Information Microarray data sets generated for this study have been deposited in the Gene Expression Omnibus under accession numbers GSE59870, GSE59872 and GSE65411. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.B.-A. (Bentires@fmi.ch).

METHODS

Mice. Lgr5-CreER^{T2} mice (C57BL/6) were generated and provided by B. Kinzel and J. Tchorz²⁵. Mice were backcrossed to the FVB background and used in this study. K8-CreER^{T2} mice (CD-1) were provided by C. Blanpain¹. The generation of *PIK3CA*^{H1047R} and *PIK3CA*^{WT} (pure FVB) was described previously^{16,18}. Tomato-reporter animals (C57BL/6) were provided by B. Roska. For lineage-tracing studies, mice of a mixed background (FVB/CD-1/C57BL/6 and FVB/C57BL/6) were used. For studies without a Tomato reporter, mice with a pure FVB background (Lgr5-CreER^{T2} model) or mixed FVB/CD-1 background (K8-CreER^{T2} model) were used. Mouse colonies were maintained in the animal facility of the Friedrich Miescher Institute for Biomedical Research and experiments were carried out in accordance with Swiss national guidelines on animal welfare and the regulations of the canton of Basel-Stadt, Switzerland.

Targeting Tomato and/or GFP expression. Adult 7- to 8-week-old female mice were induced by intraperitoneal injection of tamoxifen (Sigma; 2 mg per 25 g of body weight) for three consecutive days (diluted in sunflower seed oil, Sigma) to activate cell-specific expression of the Cre recombinase and thus expression of Tomato and/or the *PIK3CA*^{H1047R} or *WT* transgene. Tomato and/or GFP were expressed in recombined cells derived from Lgr5-positive or K8-positive cells, respectively. For the Lgr5 model, after Cre induction, GFP expression was derived from both Lgr5 promoter activity and *PIK3CA*^{H1047R} expression.

Histology and immunostaining. Tumours and dissected mammary glands were spread on a glass slide and fixed in 20% formalin for 24 h at 4 °C. Samples were then processed, embedded in paraffin and sectioned (3 µm). Immunofluorescence studies were performed using a Ventana DiscoveryUltra instrument (Roche Diagnostics) following the RUO Discovery Universal method. Briefly, slides were pretreated with CCI for 40 min and then incubated with primary antibodies for 1 h at 37 °C. After washing, secondary antibodies were incubated for 32 min at 37 °C. Slides were then washed with reaction buffer (three times), PBS (two times) and then incubated with DAPI (1 µg ml⁻¹) for 5 min. Finally, slides were rinsed with PBS (three times) and mounted in Mount Fluor (ProTaq). The following antibodies were used: anti-keratin 8/18 (guinea-pig, 1:200, Fitzgerald, 20R-CP004), anti-RFP (rabbit, 1:500, Rockland, 600-401379), anti-keratin 14 (chicken, 1:500, Covance, sig3476), anti-p63 (mouse, 1:500, Thermo Scientific, ma121871), anti-keratin 5 (rabbit, 1:500, Abcam, ab52635), anti-chicken Alexa Fluor 488, anti-chicken Alexa Fluor 568, anti-guinea pig Alexa Fluor 488, anti-mouse Alexa Fluor 488, anti-rabbit Alexa Fluor 647. All Alexa secondary antibodies were obtained from Molecular Probes (Invitrogen). Immunohistochemistry experiments were performed for keratin 14, keratin 8/18, GFP, and p63 using a Ventana DiscoveryXT instrument (Roche Diagnostics) following the Research IHC DAB Map XT procedure. Slides were treated with a mild CCI and incubated with the primary antibodies for 1 h at 37 °C. After brief washes, biotinylated donkey-anti-rabbit and biotinylated anti-guinea-pig, respectively, were applied for 32 min at 37 °C. For mouse-anti p63 detection, a monoclonal rabbit-anti-mouse antibody was applied for 32 min at 37 °C, followed by incubation with a polymer anti-rabbit conjugated with horseradish peroxidase (HRP) (ImmPRESS anti-rabbit peroxidase, Vector Laboratories) for 32 min at 37 °C. Finally, sections were counterstained with haematoxylin II and bluing reagent (4 min). Staining against ERα, SMA and keratin 5 was performed by deparaffinization followed by antigen retrieval with citrate buffer and quenching with PBS plus 3% H₂O₂. Slides were then blocked with PBS plus 2.5% normal goat serum and primary antibodies incubated overnight at 4 °C in PBS plus 1% BSA plus 0.5% Tween-20. After brief washes, secondary antibodies were then incubated in PBS plus 1% BSA for 30 min at room temperature. Signals were enhanced using the Vectastain ABC system and visualized with 3,3'-diaminobenzidine (DAB; Sigma). Haematoxylin was used as counterstain. Anti-PR staining was performed without antigen retrieval. The following antibodies were used: anti-keratin 8/18 (guinea-pig, 1:500, Fitzgerald, 20R-CP004), anti-keratin 14 (rabbit, 1:500, Thermo Scientific, Rb9020), anti-p63 (mouse, 1:1,000, Thermo Scientific, ma121871), anti-keratin 5 (rabbit, 1:1,000, Abcam, ab52635), anti-SMA (rabbit, 1:500, Thermo Scientific, Rb9010), anti-ERα (rabbit, 1:1,000, Santa Cruz, sc-542), anti-PR (rabbit, 1:200, Thermo Scientific, Rm9102), anti-GFP (rabbit, 1:50, Invitrogen, A11122). Secondary antibodies were biotinylated anti-rabbit IgG (1:200, Jackson ImmunoResearch), biotinylated anti-guinea pig IgG (1:200, Vector Laboratories) and biotinylated anti-mouse IgG (1:200, Abcam). Haematoxylin and eosin staining was performed using standard protocols.

Microscopy image acquisition. For immunofluorescence, images of stained sections were captured using a Zeiss Z1 wide-field fluorescent microscope, ×5/0.13, ×10/0.45 or ×20/0.8 (Plan-APOCHROMAT) objectives, an AxioCamMRc camera (1,024 × 1,024, pixel size 6.45 µm) and an AxioCam506 camera (2,752 × 2,208, pixel size 4.54 µm). Whole-mount fluorescent mammary gland and tumours sections were imaged on the Zeiss Axio Scan.Z1 slide scanner (ORCA-Flash4.0 camera, 2,048 × 2,048, pixel size 6.5 µm). Representative images

were cropped and processed using the ZenBlue software. For immunohistochemistry, stained sections were examined using a Nikon E600 Eclipse brightfield microscope (×20/0.5 and ×40/0.75 objectives) and images captured with a Nikon DXM1200 camera (2,592 × 1,944, pixel size 6.7 µm) using the IMS acquisition software. All images were scaled appropriately.

Quantification of immunohistochemistry. Five representative images of 8–15 tumours of each genotype were captured with a Nikon E600 Eclipse brightfield microscope (×20/0.5 objective) and positively stained tumour areas, luminal cells, and total epithelial cells quantified with Image J (Fiji). Evaluation of tissue sections was performed blindly by two independent investigators.

Quantification of double-positive tumour epithelial area. Ten tumours from each model were stained with anti-keratin 14, anti-keratin 5, anti-keratin 8/18, anti-rabbit Alexa Fluor 647, anti-chicken Alexa Fluor 568, and anti-guinea-pig Alexa Fluor 488. Whole-mount tumour sections were scanned using the Zeiss Axio Scan Z1 slide scanner. For computational reasons, the whole images of tumours were first tiled and the channels split using the ZenBlue software and Matlab. Images were then processed using Ilastik²⁶, an interactive supervised machine learning toolkit, and the subsequent prediction maps were treated with batch functions written with the Matlab programming language. For the statistical analysis shown in Fig. 4b, a total of 23 regions of tumours were subsequently tiled into 14,872 squared images of 1,024 × 1,024 pixels. In a second step, data from a set of nine selected tiles were generated by annotating a few regions representative of the predefined classes. For each of the three channels the training data were processed using the four phenotype classes: '1.background', '2.fluorescent marker (Alexa 488, 568 or 647)', '3.blood cells', and '4.Stroma'. After annotation of the different pixels of the different classes by brush stroke, features of the labelled pixels and their local neighbourhood are used to train a Random Forest classifier. In a third step, the inferred classifier was used to predict all the tiles in a batch process. The last step implemented in Matlab determined the masks for all the channels and tiles. To perform the analysis on relevant tiles, that is, tiles sufficiently covered by tissue, we performed a *k*-means clustering on a set of statistical features extracted from the tiles. This enabled us to suppress tiles that were merely in the background or those that contained non-epithelial structures. For each tile, the fluorescent marker mask corresponded to pixels having a probability above 50% in the class 'fluorescent marker (Alexa)'. In addition, a non-tissue mask was calculated by the mean of morphological filters on the union of background, blood cells and stroma classes. The mask for tissue represented the complement image of the non-tissue mask. We calculated on each tile the ratio between the areas of pixels in the mask of the fluorescent marker and those in the tissue mask. Finally, the distribution of double-labelled fluorescent marker within the whole tumour epithelial area was represented in Fig. 4b.

Preparation of mammary single-cell suspensions and labelling. Mammary glands were dissected and intra-mammary lymph nodes removed. To obtain mammary organoids, mammary glands were processed as described previously^{20,21}. To obtain single mammary epithelial cells, organoids were washed in serum-free Leibowitz L15-medium (Gibco) and digested with Hyclone HyQTase (Thermo Scientific). Single cells were washed and filtered through a 40-µm cell strainer (BD Falcon) and counted; 10⁶ cells per ml were stained with the following antibodies: PE-Cy7-CD45 (Biolegend; clone 30-F11), APC-Sca1 (Biolegend; clone E13-161.7), PerCP-Cy5.5-CD24 (Biolegend; clone M1/69), PE-CD49f (BD-Pharmingen), Alexa700-CD24 (Novus Biologicals; clone M1/69) and DAPI (2 µg ml⁻¹, Invitrogen).

Flow cytometry. FACS was carried out with a BD FACSaria III (Becton Dickinson) using a 100-µm nozzle. Cells were gated based on their forward- and sideward-scatter. Pulse-width was used to exclude doublets. DAPI-negative/CD45-negative cells were gated for Tomato or GFP. CD24, Sca1 and CD49f subsets were then gated on GFP-positive epithelium. Tomato-positive cells were gated using only CD24, CD45 and Sca1 antigens. The same numbers of living cells were recorded in each condition. FACS data were analysed using FlowJo (Tree Star). Total cell numbers were determined by enumerating the total epithelial content after single-cell isolation and calculating back based on the percentages obtained from antibody staining, sorting and FlowJo analysis. Cell numbers were subsequently normalized to one animal (3–9 independent FACS experiments of 1–5 pooled animals per time point were quantified; ± s.e.m.; *P* < 0.05, Student's *t*-test).

In vitro colony formation assay and quantification. Freshly sorted cells of each subpopulation (500 cells) were plated as previously described²¹. Seven days later, the colonies were fixed with acetone/methanol (1:1), washed, blocked with 2.5% normal goat serum and stained with anti-keratin 8/18 (guinea pig, 1:500, Fitzgerald, 20R-CP004), anti-keratin 14 (rabbit, 1:500, Thermo Scientific, Rb9020), DAPI (2 µg ml⁻¹, Invitrogen), anti-guinea-pig Alexa Fluor 488 and anti-rabbit Alexa Fluor 647 (1:1,000, Invitrogen). Colonies were imaged using the Zeiss Z1 wide-field fluorescent microscope (×5/0.13 DIC). Colonies were

defined as a cluster of more than five cells. The number of colonies per well was determined manually. Colonies containing more than 20% of keratin 8/18/keratin 14 double-positive cells were defined as 'double-positive'.

In vitro mammosphere culture. Mammosphere cultures were performed as described previously²⁷. Freshly sorted subsets (luminal CD24^{Hi}Sca1⁻ and CD24^{Hi}Sca1⁺) from adult FVB and uninduced *PIK3CA*^{H1047R} females were plated at 20,000 cells per ml in 6-well ultra-low attachment plates (Falcon) in DMEM/F12 medium (Gibco) supplemented with 5 µg ml⁻¹ insulin, 0.5 µg ml⁻¹ hydrocortisone, 2% B27 (Invitrogen), 20 ng ml⁻¹ EGF and bFGF (BD Biosciences) and cholera toxin (Sigma), and cultured at 37 °C in 5% CO₂. After 5 days, mammospheres were collected and dissociated into single cells using HYQ-tase (Gibco) and counted. Luminal subsets from FVB control and *PIK3CA*^{H1047R} animals were subsequently treated with TAT-Cre (Millipore) (0.5 µM) at a density of 20,000 cells per ml in 4 ml overnight at 37 °C in 5% CO₂. The next morning, the medium was replaced and cells were cultured for 72 h at 37 °C in 5% CO₂ to allow for maximal recombination and expression of *PIK3CA*^{H1047R}. After 72 h, *PIK3CA*^{H1047R}-GFP^{+/+} cells were sorted and plated into 24-well ultra-low attachment plates in medium (described earlier) supplemented with 2% Matrigel (growth factor reduced; BD, 356230) at a density of 1,000 cells per well. Control and *PIK3CA*^{H1047R} spheres from each subset were enumerated every 7 days, at which point spheres were dissociated and re-plated at a density of 1,000 cells per well.

Mammary fat pad transplantation. For limiting dilution transplantation, freshly sorted cells from *Lgr5*-CreER^{T2} control and *Lgr5*-CreER^{T2}/*PIK3CA*^{H1047R} animals (pure FVB background) and from K8-CreER^{T2} control and K8-CreER^{T2}/*PIK3CA*^{H1047R} F1-hybrid (FVB/CD-1) littermates were used. Sorted cells were resuspended in limiting dilution numbers in PBS plus 2% FCS with 25% Matrigel (growth factor reduced; BD, 356230) and injected in 20-µl volumes into inguinal glands of 3-week-old FVB females that had been cleared of endogenous mammary epithelium²⁸. Cells from control and from *PIK3CA*^{H1047R} animals were injected in the same animal on opposite sides. After 8 weeks, glands of the recipients were removed for evaluation. Glands were spread on a glass slide and fixed in Carnoy's fixative overnight. Whole-mount staining with carmine alum was performed as previously described²¹ and scanned with an Epson 1600 Pro scanner. An outgrowth was defined as an epithelial structure composed of ducts arising from a central point with lobules and/or terminal end buds⁸. Frequencies of mammary-repopulating units between different cell populations were calculated and statistically compared using the Extreme Limiting Dilution Analysis (ELDA)²⁹ online tool (<http://bioinf.wehi.edu.au/software/elda/>).

Immunoblotting. Lysates from mammary glands were prepared by lysing cryo-homogenized mammary gland powder in RIPA buffer (50 mM Tris-HCl pH 8, 150 mM NaCl, 1% NP-40, 0.5% sodium deoxycholate, 0.1% SDS) supplemented with 1× protease inhibitor cocktail (Complete Mini, Roche), 0.2 mM sodium orthovanadate, 20 mM sodium fluoride and 1 mM phenylmethylsulfonyl fluoride. Lysates (30–80 µg) were subjected to SDS-PAGE, transferred to PVDF membranes (Immobilon-P, Millipore) and blocked for 1 h at room temperature with 5% milk or BSA in PBS/0.05% Tween 20. Membranes were then incubated overnight with primary antibodies (1:200–1:3,000) and exposed to secondary HRP-coupled anti-mouse or anti-rabbit antibodies at 1:5,000–10,000 for 1 h at room temperature. Results are representative of at least three different experiments. The following antibodies were used: anti-AKT pan (Cell Signaling), anti-pAKT (Ser473, Cell Signaling), anti-p110α (Cell Signaling), anti-ERK2 (Santa Cruz) and anti-keratin pan (SantaCruz). Blot densities were quantified using ImageJ and normalized to pan-keratin for epithelial content.

RNA isolation. For pre-neoplastic gene expression profiling, mammary epithelial subsets were FACS sorted as described earlier into extraction buffer and total RNA from 250 or 2,000 sorted cells was isolated using an Arcturus PicoPure RNA Isolation Kit (Life Technologies). Subsets of pooled mammary glands of 2–3 oestrus-synchronized animals (confirmed by vaginal smear) per genotype in three independent sortings were collected for microarray analysis. RNA from mammary subsets of four animals per genotype was collected for qRT-PCR. For tumour gene expression profiling and expression of *Lgr5* in the mammary gland, total RNA from 50 mg cryo-homogenized tumour tissue or nipple and distal area of mammary glands were extracted using the TRIzol method (Life Technologies) according to the manufacturer's instructions. Genomic DNA was removed by DNase I digestion (Qiagen) and RNA purified using an RNeasy Plus Mini Kit (Qiagen). RNA concentration was measured with a Nanodrop 1000 machine and RNA quality assessed using an Agilent 2100 bioanalyzer and RNA Pico Chips or RNA Nano Chips.

Microarray. Total RNA from 250 sorted cells (pre-neoplastic *Lgr5*), 2,000 sorted cells (pre-neoplastic K8) or 100 ng from tumour tissue was used as the input for synthesis of amplified cDNA with the NuGen Ovation Pico WTA System (NuGen). The resulting double-stranded cDNA was fragmented and labelled

using the Affymetrix GeneChip WT Terminal Labelling kit (Affymetrix). Affymetrix Gene Chip Mouse gene 1.0 ST microarrays were hybridized according to the GeneChip Whole Transcript (WT) Sense Target Labelling Assay Manual (Affymetrix) with a hybridization time of 16 h. Scanning was performed with Affymetrix GCC Scan Control Software v. 3.0.0.1214 on a GeneChip Scanner 3000 7 G with autoloader.

Normalization and analysis of mammary epithelial subset microarray data. The arrays for Gene Expression Omnibus accession numbers GSE40875 (ref. 21), GSE59870 (Fig. 2a) and GSE65411 (Fig. 2b) data sets were RMA normalized using the bioconductor package *affy* (R3.0.1/Bioconductor 2.13). Heat maps with unscaled normalized expression values for selected genes were plotted with the *heatmap2* function of the *gplots* package. Differential gene expression for the comparison of mature luminal (LM) or luminal progenitor (LP) versus myoepithelial (MYO) cells was calculated with *limma*. The function *topTable* was used to select the top 300 upregulated genes as luminal mature versus myoepithelial UP (Luminal_Mature UP, MEIER-ABT) and luminal progenitors versus myoepithelial UP (Luminal_progenitors UP, MEIER-ABT), and the top 300 upregulated genes as myoepithelial versus luminal mature UP (Myoepithelial UP, MEIER-ABT). Signatures LIM_MAMMARY_LUMINAL_MATURE_DN, LIM_MAMMARY_LUMINAL_MATURE_UP, LIM_MAMMARY_LUMINAL_PROGENITOR_DN, LIM_MAMMARY_LUMINAL_PROGENITOR_UP, LIM_MAMMARY_STEM_CELL_DN and LIM_MAMMARY_STEM_CELL_UP were downloaded from <http://www.broadinstitute.org/gsea/msigdb/search.jsp> in gmt format, combined with the signatures described earlier, and used in the PGSEA and *smcPlot* functions of the Bioconductor PGSEA package.

qRT-PCR. RNA was converted into cDNA using SuperScript III Reverse Transcriptase (Invitrogen). Quantitative real-time PCR was performed on unamplified cDNA normalized to the number of sorted cells for each subset (6,000–10,000 sorted cells). Taqman probes (Life Technologies) and Taqman Universal PCR Mastermix (Applied Biosystems) were applied. The following Taqman probe identifiers were used: *Krt14* Mm00516879_m1, *Krt5* Mm01305291_g1, *Lgr5* Mm00438890_m1, *Vim* Mm01333430_m1, *Krt8* Mm00835759_m1, *Krt18* Mm01601702_g1, *Gata3* Mm00484683_m1, *Elf5* Mm00468732_m1, *Csn2* Mm04207885_m1 and *Axin2* Mm00443610_m1. Cycling was performed with StepOne Plus Real-time PCR Systems (Applied Biosystems). The results are representative of three qRT-PCR experiments of pooled mammary subsets from four animals of each genotype. Results of *Csn2* (CD24^{Hi}Sca1^{-/+}) and *Krt14* (CD24^{Hi}Sca1⁺) from K8-CreER^{T2}/*PIK3CA*^{H1047R} cells and of *Krt5* and *Lgr5* (CD24^{Hi}Sca1⁻) from *Lgr5*-CreER^{T2}/*PIK3CA*^{H1047R} cells are representative of two experiments. Statistical data analysis was performed using $\Delta\Delta C_T$ values.

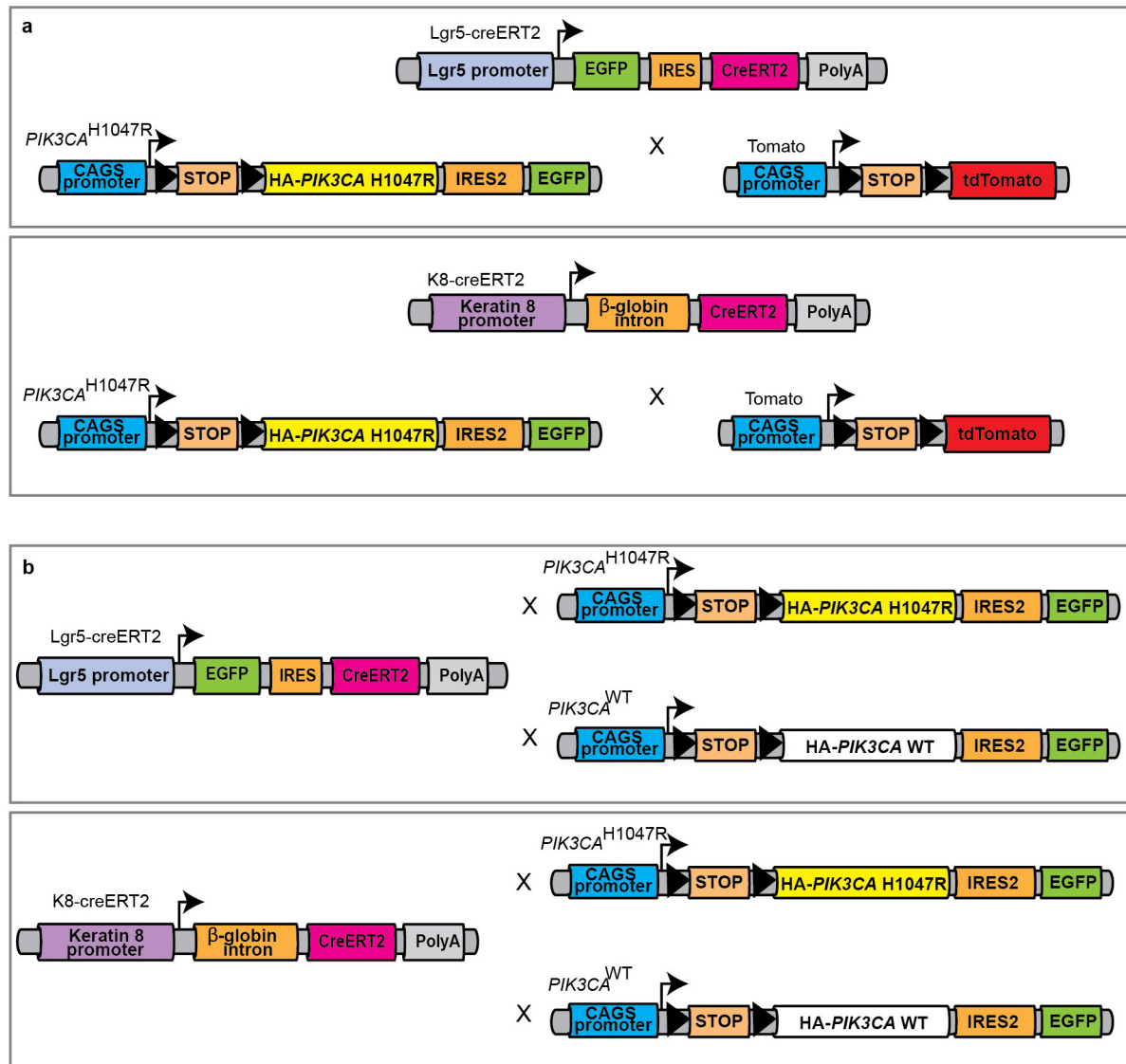
Normalization and analysis of tumour mouse microarray data. Mouse Affymetrix microarrays (Gene Expression Omnibus accession number GSE59872) were background corrected, quantile normalized, and log₂ transcript cluster expression values were calculated using the *rma()* function of the Bioconductor package "oligo"³⁰. Transcript cluster identifiers were mapped to Entrez Gene identifiers using the Bioconductor package "mogene10sttranscriptcluster.db", and transcript clusters associated with none or multiple genes were removed. Where multiple transcript clusters were associated with a single gene that with the maximal variance across samples was selected; this resulted in a total number of 20,365 transcript clusters with unique gene assignments. Principal component analysis was performed on log₂ expression values from which the mean over samples had been subtracted for each gene. Hierarchical clustering of samples was performed using $(1 - r)$ as distance metric, with r being the Pearson's correlation coefficient for log₂ expression values between each pair of samples. The clustering dendrogram was visualized using the "dendextend" R package.

Analysis of human TCGA breast cancer data and comparison to mouse. Human breast cancer expression data and corresponding clinical data⁵ were obtained from https://tcga-data.nci.nih.gov/docs/publications/brca_2012/, corresponding to the 11 November, 2011 data freeze that contains 522 tumour samples with clinical annotation. Human gene symbols were mapped to Entrez Gene identifiers and genes selected that are one-to-one homologues between human and mouse according to Homologene (build 68, downloaded from <ftp://ftp.ncbi.nlm.nih.gov/pub/HomoloGene/build68/31/>) and that were measured on both human and mouse experimental platforms; this resulted in 13,969 genes. To reduce technical differences between human and mouse samples, all samples were scaled to a standard deviation of 1, and the within-species mean was subtracted from each gene. Heat map and clustering of the combined human and mouse expression data were performed using the function *heatmap()* from the R package "NMF"³² on the top 1,000 genes ranked by variance across human samples. Mouse samples were compared with human PAM50 tumour subtypes (Normal-like, Luminal A, Luminal B, HER2-enriched and Basal-like) by calculating Pearson's correlation coefficients for each mouse sample against the averages of all human samples within each subtype.

Survival analysis. Tumour incidence was determined by palpation. Kaplan–Meier plots were generated using the survival calculation tool from Graphpad Prism and significance was calculated using the log-rank test.

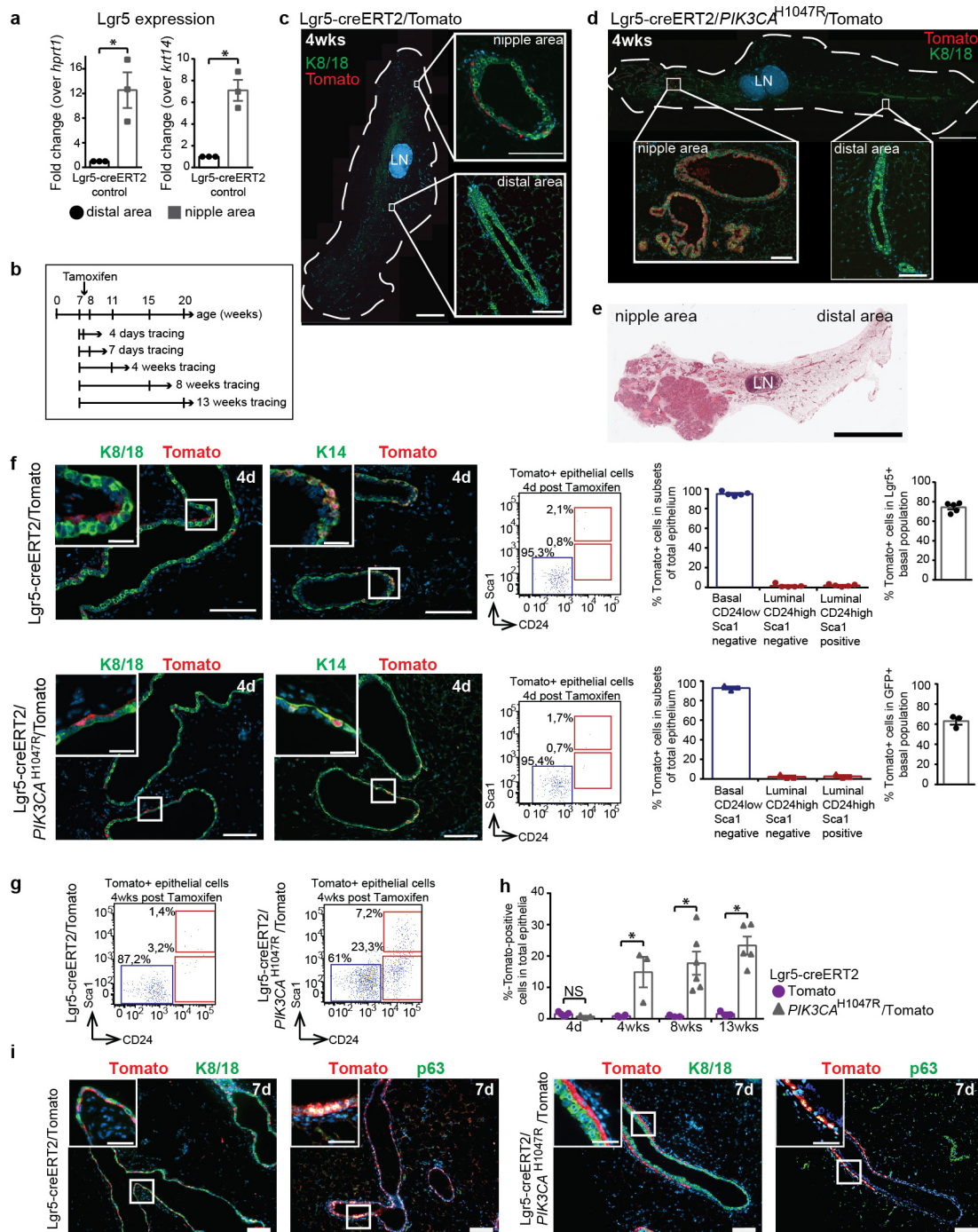
Statistical data analysis. The number of mice was calculated by performing power analysis using data from small pilot experiments. Values represent the means \pm s.e.m. or \pm s.d. Depending on the type of experiments, data were tested using unpaired Student's *t*-test or log-rank test. * $P < 0.05$ was considered statistically significant. The experiments were not randomized.

25. Kinzel, B. *et al.* Functional roles of Lgr4 and Lgr5 in embryonic gut, kidney and skin development in mice. *Dev. Biol.* **390**, 181–190 (2014).
26. Sommer, C., Strähle, C., Köthe, U. & Hamprecht, F. A. Ilastik: Interactive Learning and Segmentation Toolkit. *Proc. Eighth IEEE Int. Symp. Biomed. Imag.* 230–233 (2011).
27. Cicalese, A. *et al.* The tumor suppressor p53 regulates polarity of self-renewing divisions in mammary stem cells. *Cell* **138**, 1083–1095 (2009).
28. Deome, K. B., Faulkin, L. J. Jr, Bern, H. A. & Blair, P. B. Development of mammary tumors from hyperplastic alveolar nodules transplanted into gland-free mammary fat pads of female C3H mice. *Cancer Res.* **19**, 515–520 (1959).
29. Hu, Y. & Smyth, G. K. ELDA: extreme limiting dilution analysis for comparing depleted and enriched populations in stem cell and other assays. *J. Immunol. Methods* **347**, 70–78 (2009).
30. Carvalho, B. S. & Irizarry, R. A. A framework for oligonucleotide microarray preprocessing. *Bioinformatics* **26**, 2363–2367 (2010).
31. NCBI Resource Coordinators. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **42**, D7–D17 (2014).
32. Gaujoux, R. & Seoighe, C. A flexible R package for nonnegative matrix factorization. *BMC Bioinformatics* **11**, 367 (2010).



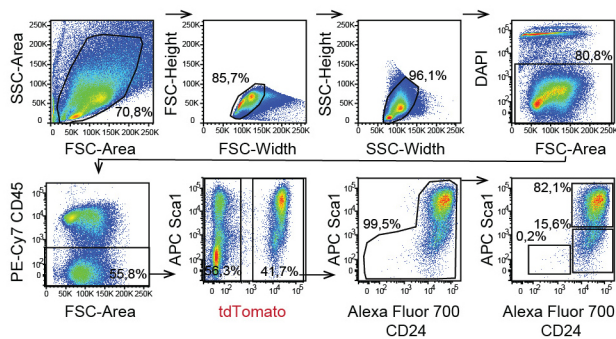
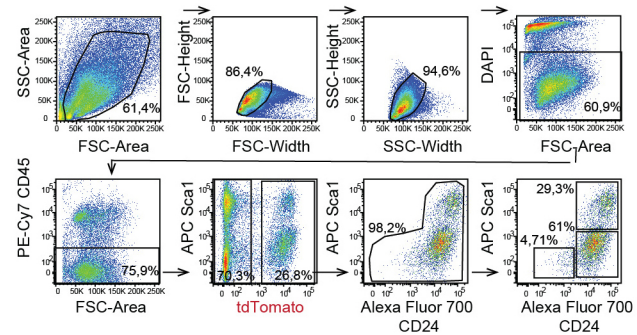
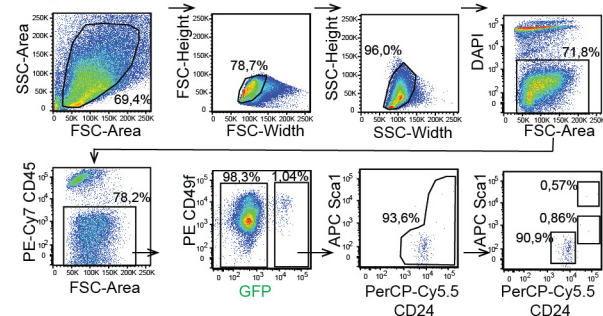
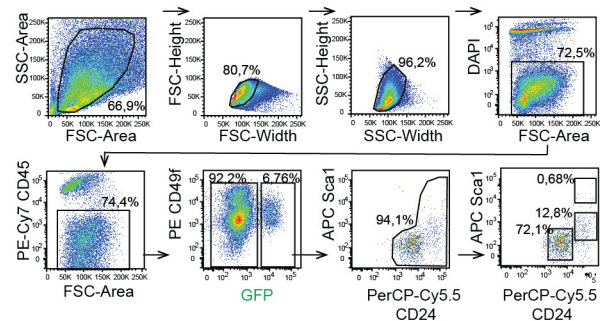
Extended Data Figure 1 | Scheme depicting mouse lines generated for lineage-tracing studies. **a**, Lgr5-CreER^{T2} (ref. 25) or K8-CreER^{T2} (ref. 1) animals were crossed to transgenic lox-STOP-lox *PIK3CA*^{H1047R} (ref. 16) and/or Tomato-reporter mice, generating Lgr5-CreER^{T2}/Tomato, K8-CreER^{T2}/Tomato, Lgr5-CreER^{T2}/*PIK3CA*^{H1047R}/Tomato and K8-CreER^{T2}/*PIK3CA*^{H1047R}/Tomato animals for lineage-tracing studies. Lgr5-CreER^{T2}/Tomato and

K8-CreER^{T2}/Tomato animals were used as controls. **b**, Lgr5-CreER^{T2} (ref. 25) and K8-CreER^{T2} (ref. 1) animals were crossed to lox-STOP-lox *PIK3CA*^{H1047R} (ref. 16) or *PIK3CA*^{WT} (ref. 18) animals. Lgr5-CreER^{T2} and K8-CreER^{T2} animals were used as controls. Tamoxifen injection induces *PIK3CA*^{H1047R}, *PIK3CA*^{WT} and/or Tomato expression.



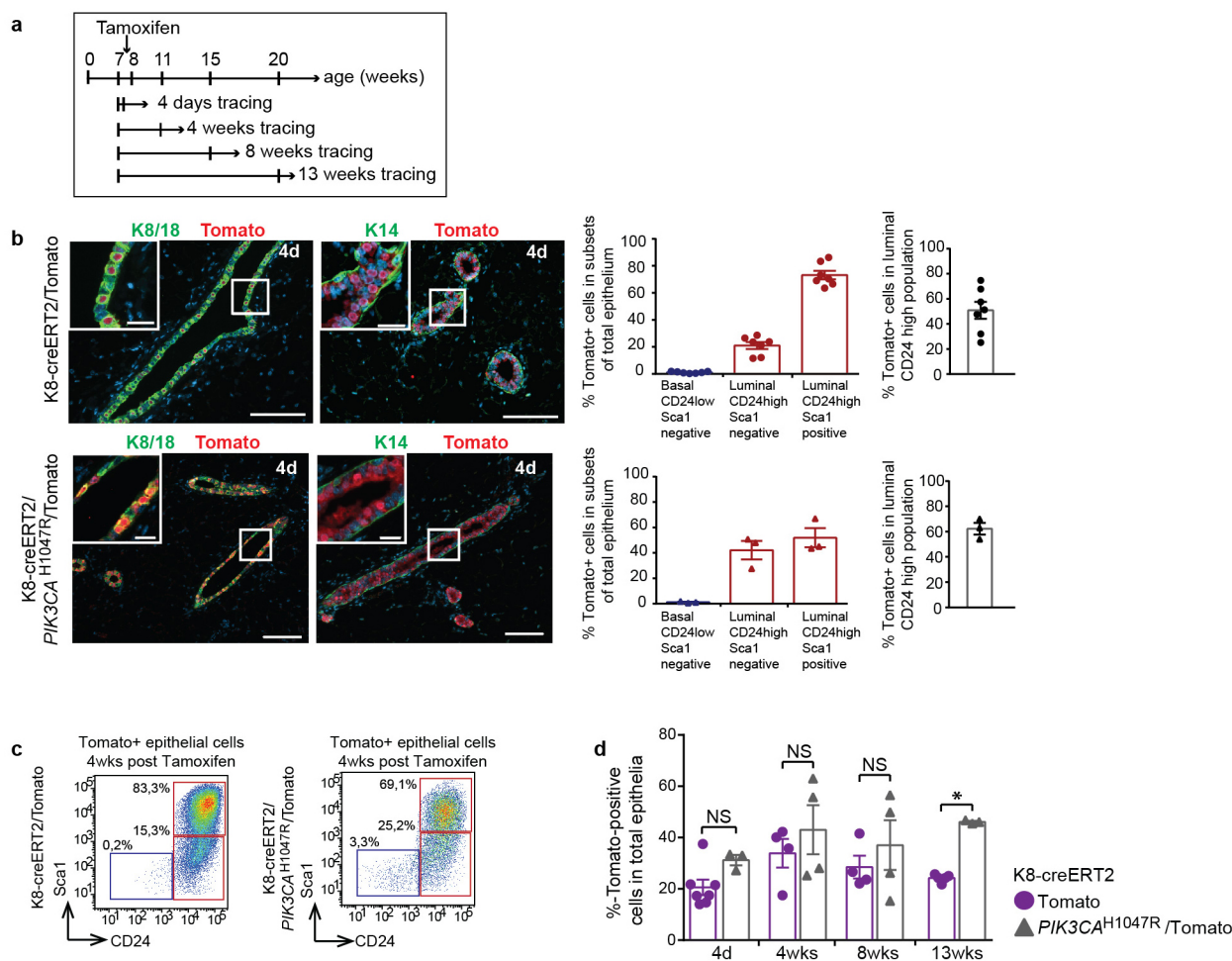
Extended Data Figure 2 | Lgr5-CreER^{T2}/Tomato and PIK3CA^{H1047R}/Tomato labelling in the mammary nipple area. **a**, Lgr5 expression in the nipple and distal area of Lgr5-CreER^{T2} glands ($n = 3$ mice). **b**, Tracing scheme. **c**, **d**, Representative images of mammary glands after 4 weeks tracing ($n = 3$ mice for each genotype). Scale bars, 2 mm, 100 μ m (magnifications). **e**, Representative haematoxylin and eosin staining of an Lgr5-CreER^{T2}/PIK3CA^{H1047R} mammary gland with a tumour. Scale bar, 500 μ m. LN, lymph node. **f**, Representative images, FACS plots and quantification of 4 days tracing (24 h after the last tamoxifen injection) (top: immunofluorescence: $n = 3$ mice; FACS: $n = 5$ technical replicates (each 1–2 pooled mice); bottom:

immunofluorescence: $n = 3$ mice; FACS: $n = 3$ technical replicates (each 1 mouse)). Scale bars, 100 μ m; 20 μ m (magnifications). **g**, Representative FACS plots of 4-week tracing. **h**, Percentage of total Tomato-positive cells in the tracing experiments (Lgr5-CreER^{T2}/Tomato: 4 days $n = 5$, 4 weeks $n = 4$, 8 and 13 weeks $n = 3$ technical replicates (each 1–2 pooled mice); Lgr5-CreER^{T2}/PIK3CA^{H1047R}/Tomato: 4 days $n = 3$, 4 weeks $n = 3$, 8 weeks $n = 6$ and 13 weeks $n = 5$ technical replicates (each 1–2 pooled mice)). **i**, Representative images of 7 days tracing (left $n = 4$ mice; right $n = 2$ mice). Scale bars, 100 μ m, 50 μ m (magnifications). Bar graphs show means \pm s.e.m.; two-sided unpaired Student's t -test; * $P < 0.05$; NS, not significant.

a K8-creERT2/Tomato**b** K8-creERT2/*PIK3CA*^{H1047R}/Tomato**c** Lgr5-creERT2/*PIK3CA*^{WT}**d** Lgr5-creERT2/*PIK3CA*^{H1047R}**Extended Data Figure 3 | Gating scheme for FACS experiments.**

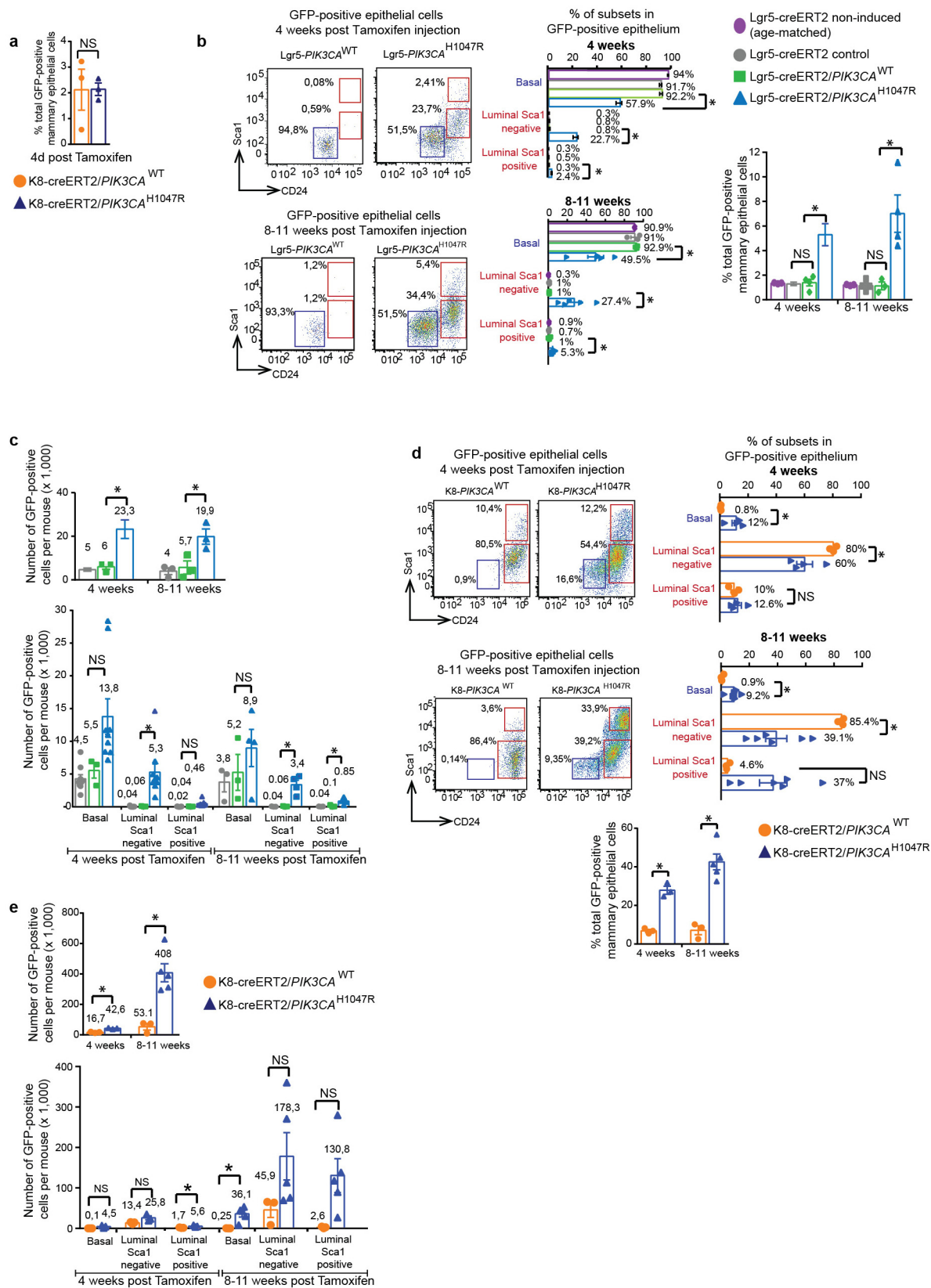
a–d, Representative FACS plots of K8-CreER^{T2}/Tomato (**a**), K8-CreER^{T2}/*PIK3CA*^{H1047R}/Tomato (**b**), Lgr5-CreER^{T2}/*PIK3CA*^{WT} (**c**) and Lgr5-CreER^{T2}/*PIK3CA*^{H1047R} (**d**) animals 4 weeks after tamoxifen injection. The gating

strategy shown illustrates the elimination of doublets, dead cells (DAPI^{Hi}), and white blood cells (CD45⁺) and the sorting of Tomato- or GFP-positive mammary epithelial subsets (basal CD24^{Lo}Sca1[−], luminal CD24^{Hi}Sca1⁺).



Extended Data Figure 4 | K8-CreERT2/Tomato and $PIK3CA^{H1047R}$ /Tomato labelling in the mammary gland. **a**, Scheme depicting timeline of tracing experiments. **b**, Representative images and FACS quantifications of K8-CreERT2/Tomato and K8-CreERT2/ $PIK3CA^{H1047R}$ /Tomato mammary glands 4 days after tamoxifen (24 h after the last tamoxifen injection) (top: immunofluorescence: $n = 5$ mice; FACS: $n = 7$ technical replicates (each 1–2 pooled mice)); bottom: immunofluorescence $n = 3$ mice; FACS: $n = 3$ technical

replicates (each 1 mouse)). Scale bars, 100 μ m, 20 μ m (magnifications). **c**, Representative FACS plots of 4-week Tomato tracing. **d**, Percentage of total Tomato-positive cells in mammary glands (K8-CreERT2/Tomato: 4 days $n = 7$, 4 and 8 weeks $n = 4$, 13 weeks $n = 5$ technical replicates (each 1–3 pooled mice)); K8-creERT2/ $PIK3CA^{H1047R}$ /Tomato: 4 days $n = 3$, 4 and 8 weeks $n = 4$ and 13 weeks $n = 3$ technical replicates (each 1–2 pooled mice)). Bar graphs show means \pm s.e.m. * $P < 0.05$; two-sided unpaired Student's t -test.

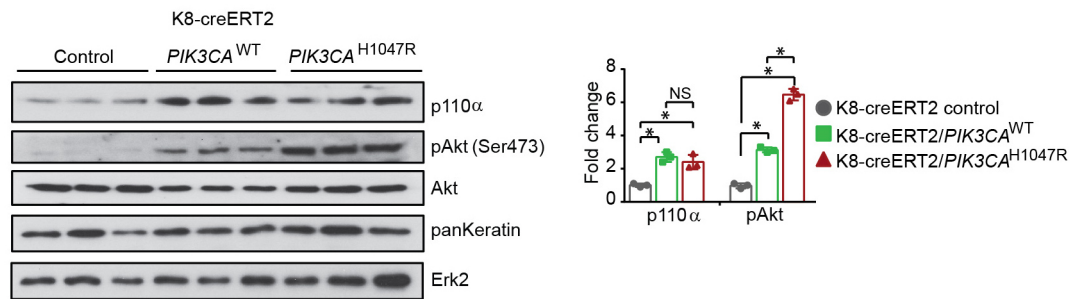


Extended Data Figure 5 | Tracing of GFP-positive mammary subsets.

a, Percentage of GFP-labelled cells in K8-CreER^{T2}/*PIK3CA*^{H1047R} versus K8-CreER^{T2}/*PIK3CA*^{WT} animals 4 days after tamoxifen (24 h after the last tamoxifen injection) ($n = 3$ technical replicates, 2 mice per genotype).

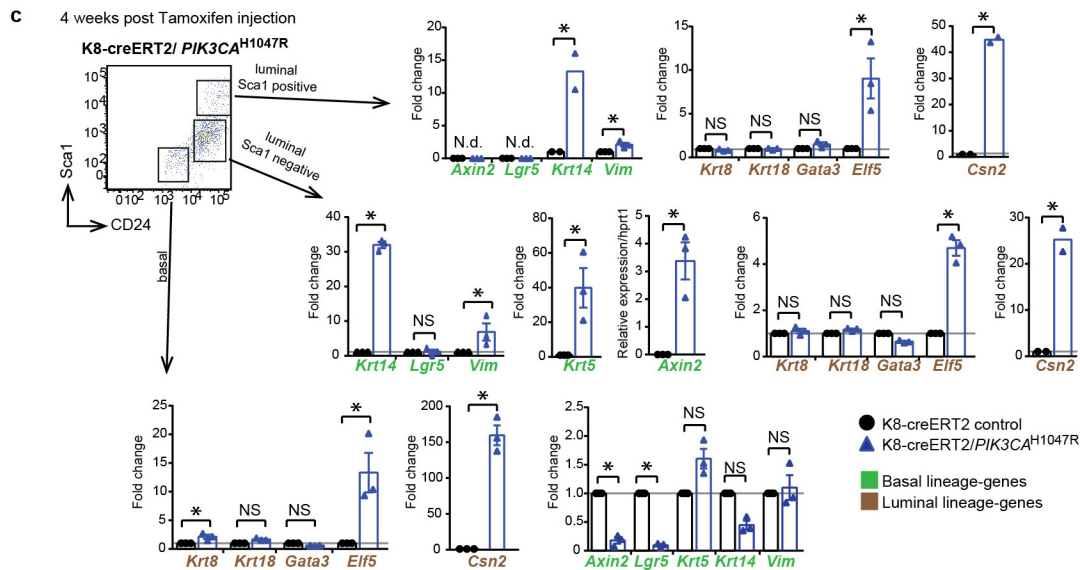
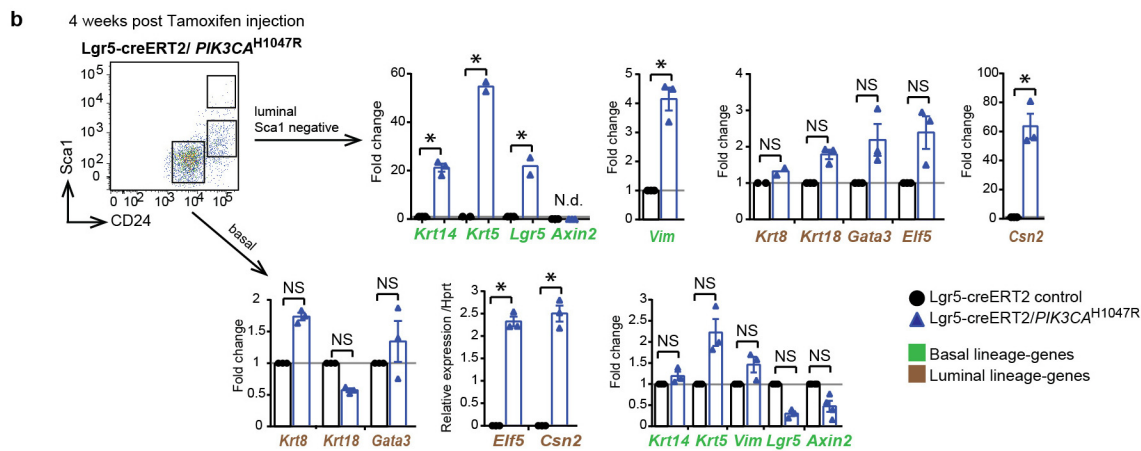
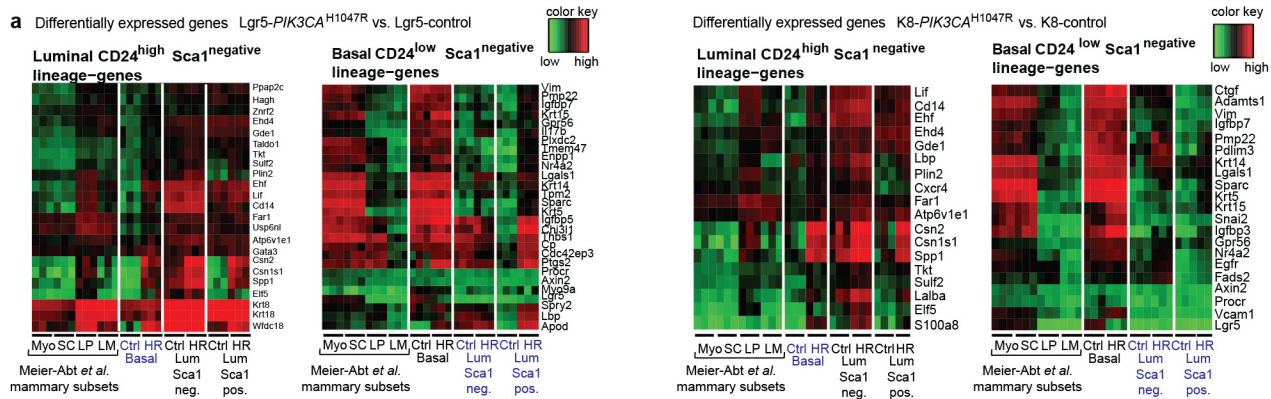
b, d, Representative FACS plots and percentages of GFP-positive cells in mammary gland subsets and total mammary epithelial cells 4 and 8–11 weeks after tamoxifen. **c, e**, Bar graphs showing total numbers of GFP-positive cells and numbers of GFP-positive cells in basal (CD24^{Lo}Sca1[−]) and luminal (CD24^{Hi}Sca1⁺) subsets of Lgr5-CreER^{T2}/*PIK3CA*^{H1047R} (**c**) and

K8-CreER^{T2}/*PIK3CA*^{H1047R} (**e**) mammary epithelial cells. **b, c**, 4 weeks: non-induced control $n = 3$, control $n = 9$, *PIK3CA*^{WT} $n = 3$, *PIK3CA*^{H1047R} $n = 9$ sortings with each 1–4 pooled mice; 8–11 weeks: non-induced control $n = 3$, control $n = 3$, *PIK3CA*^{WT} $n = 3$, *PIK3CA*^{H1047R} $n = 4$ sortings with each 1–4 pooled mice. **d, e**, 4 weeks: *PIK3CA*^{WT} and *PIK3CA*^{H1047R} $n = 3$ sortings with each 1–5 pooled mice; 8–11 weeks: *PIK3CA*^{WT} $n = 4$, *PIK3CA*^{H1047R} $n = 5–6$ sortings with each 1–4 pooled mice. Bar graphs show means \pm s.e.m.; two-sided unpaired Student's *t*-test; * $P < 0.05$; NS, not significant.



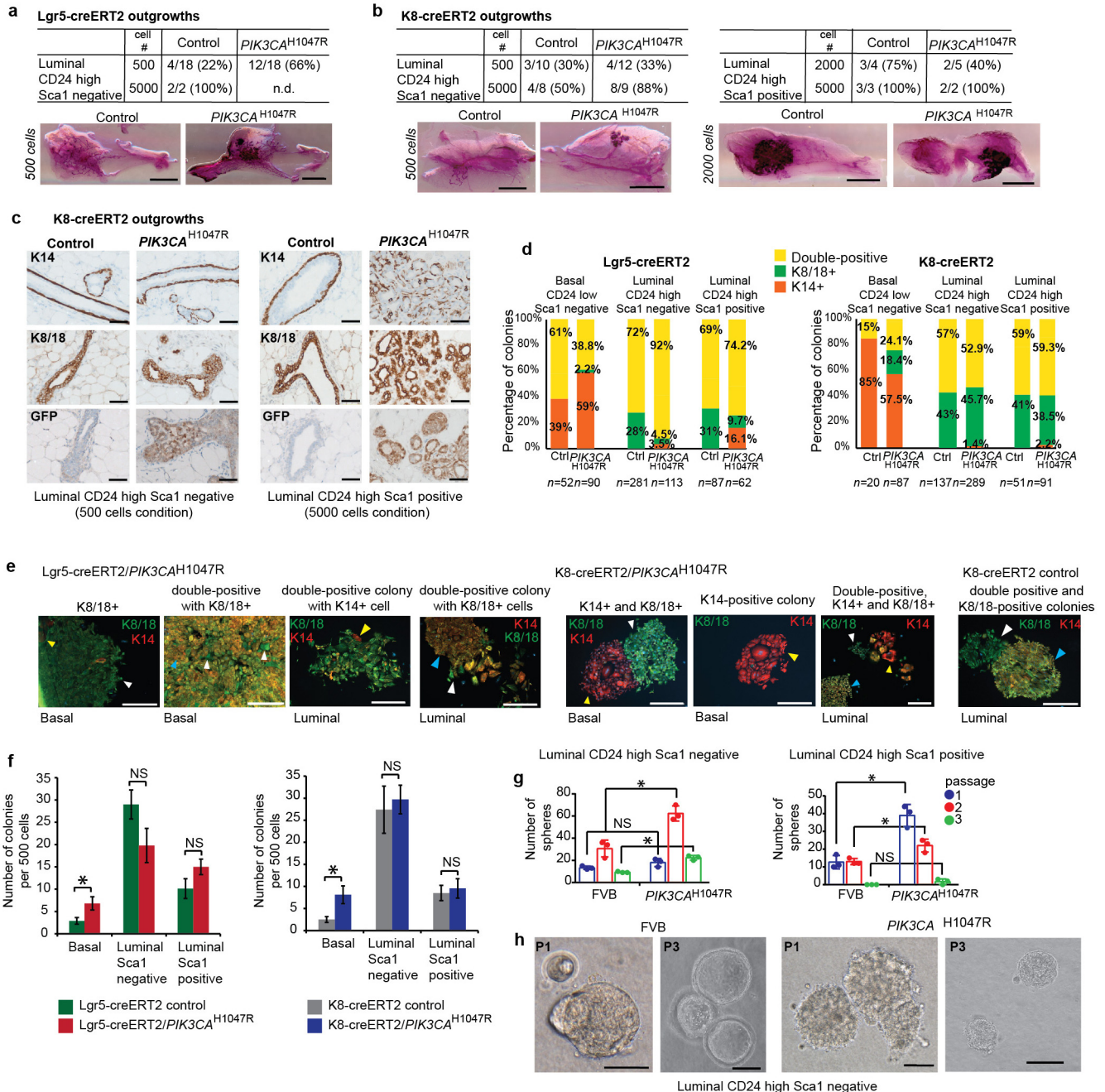
Extended Data Figure 6 | Expression of *PIK3CA*^{H1047R} induces Akt phosphorylation. Immunoblot and quantification of lysates from K8-CreER^{T2} control, *PIK3CA*^{WT} and *PIK3CA*^{H1047R} mammary glands 4 weeks after tamoxifen for p110α, pAkt, Akt, pan-keratin and Erk2 (loading control).

$n = 3$ mice per genotype. Protein levels were normalized to pan-keratin for normalization of epithelial content. Bar graphs depict fold change over control lysate. Bar graph shows means \pm s.d.; two-sided unpaired Student's *t*-test; * $P < 0.006$; NS, not significant.



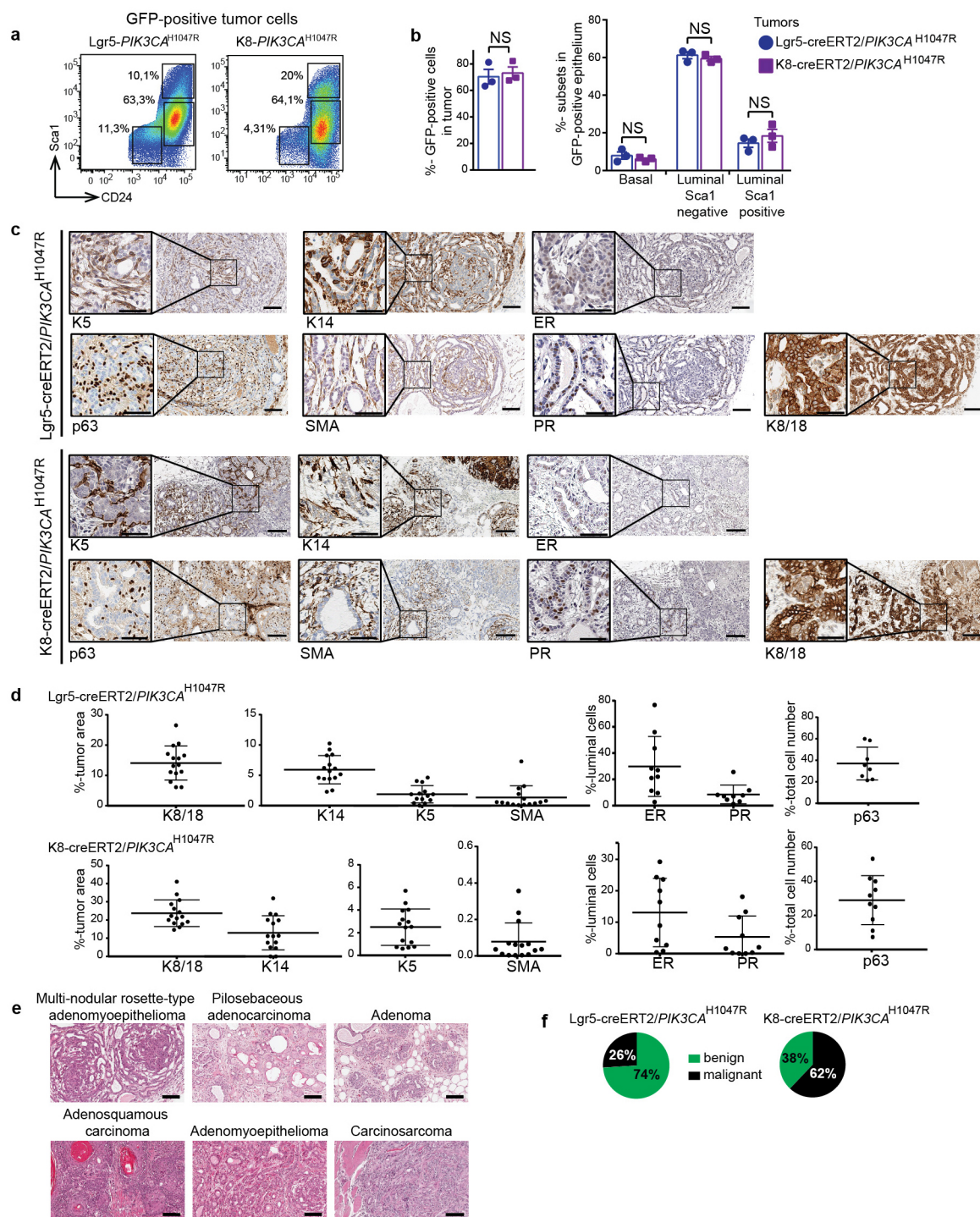
Extended Data Figure 7 | Expression of basal- and luminal-lineage genes in *PIK3CA*^{H1047R} subsets. **a**, Expression heat maps of selected luminal and basal genes (left: Lgr5-CreER^{T2}/*PIK3CA*^{H1047R} versus control; right: K8-CreER^{T2}/*PIK3CA*^{H1047R} versus control). LM, mature luminal cells; LP, luminal progenitors; Myo, myoepithelial; SC, stem-cell enriched. **b, c**, Expression profiles of basal- and luminal-lineage genes in mammary subsets of

Lgr5-CreER^{T2}/*PIK3CA*^{H1047R} compared with Lgr5-CreER^{T2} control (**b**) and K8-CreER^{T2}/*PIK3CA*^{H1047R} compared with K8-CreER^{T2} control animals (**c**). The qRT-PCR results are representative of 2–3 experiments of 4 pooled animals of each genotype. Bar graphs show means \pm s.e.m.; two-sided unpaired Student's *t*-test; **P* < 0.05; NS, not significant; N.d., not detected.



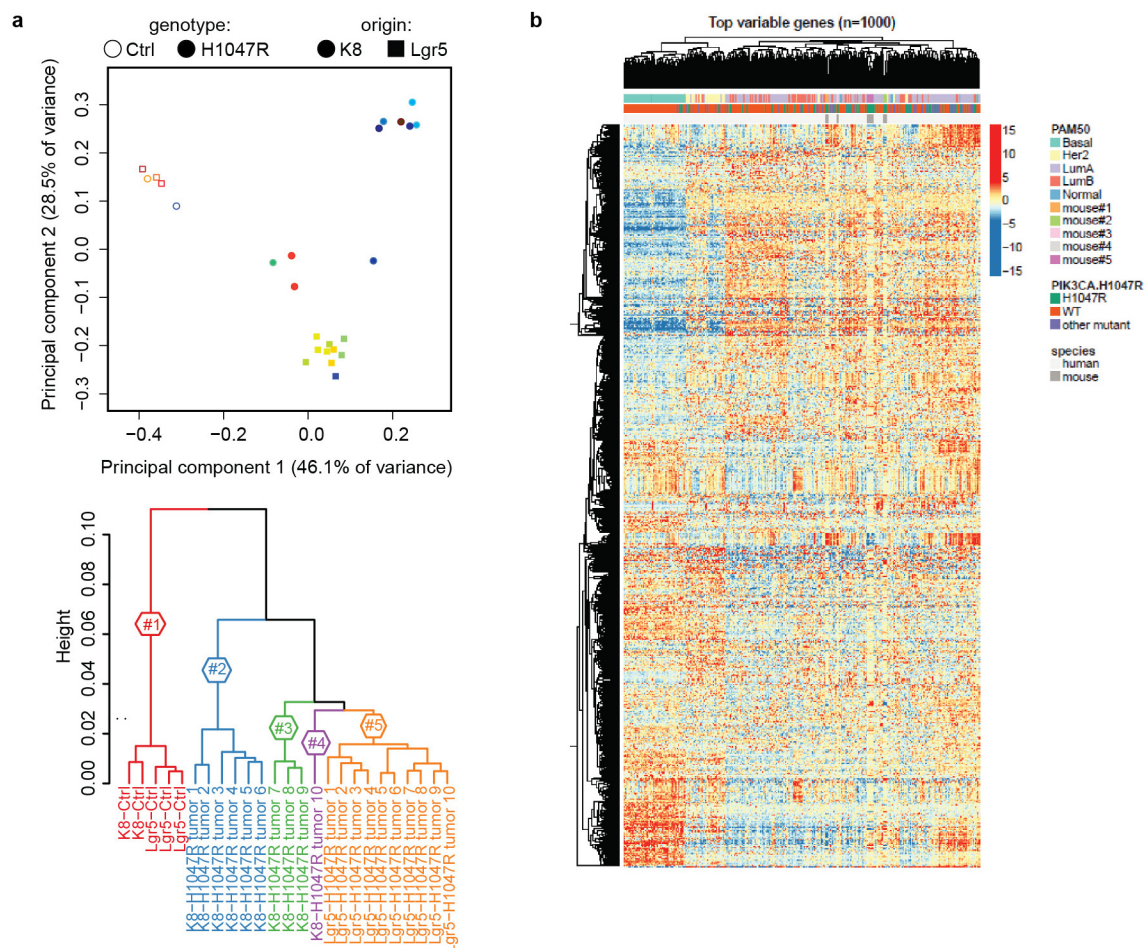
Extended Data Figure 8 | Luminal *PIK3CA*^{H1047R} cells repopulate a mammary gland. **a, b**, Number of outgrowths in cleared-fat pad transplantation of GFP-negative Lgr5-CreERT² control and GFP-positive Lgr5-CreERT²/*PIK3CA*^{H1047R}-expressing luminal subsets (CD24^{Hi}Sca1⁻) (**a**) and GFP-negative K8-CreERT² control and GFP-positive K8-CreERT²/*PIK3CA*^{H1047R}-expressing luminal subsets (left, CD24^{Hi}Sca1⁻; right, CD24^{Hi}Sca1⁺) (**b**). Representative carmine-stained whole mounts (bottom). Scale bars, 500 μ m. **c**, Representative immunostained sections. Scale bars, 50 μ m. **a–c**, Data from three independent experiments. **d**, Percentage of K14-, K8/18- and double-positive (K14/K8/18) colonies derived from Lgr5-CreERT²/*PIK3CA*^{H1047R}, Lgr5-CreERT² control (left, pooled data from $n = 4$ independent experiments (1–5 pooled mice)), K8-CreERT²/*PIK3CA*^{H1047R} and K8-CreERT² control subsets (right, pooled data from $n = 3$ independent experiments (1–5 pooled mice)). Total number of quantified colonies is shown. **e**, Representative images of colonies. Arrowheads indicate K8/18- (white), K14- (yellow) and double-positive (blue) colonies. Scale bars, 500 μ m. **f**, Number of colonies derived from basal and luminal cells from Lgr5- and K8-CreERT²/*PIK3CA*^{H1047R} and control mice. Left, pooled data from three independent

sortings (each 1–5 pooled animals), total $n = 8$ (control), $n = 10$ (mutant) technical replicates for basal subset, $n = 9$ (control), $n = 5$ (mutant) technical replicates for luminal CD24^{Hi}Sca1⁻ subset and $n = 8$ (control), $n = 4$ (mutant) technical replicates for luminal CD24^{Hi}Sca1⁺ subset. Right, pooled data from two independent sortings (each 1–5 pooled animals), total $n = 8$ (control), $n = 10$ (mutant) technical replicates for basal subset, $n = 5$ (control), $n = 10$ (mutant) technical replicates for luminal CD24^{Hi}Sca1⁻ subset and $n = 6$ (control), $n = 9$ (mutant) technical replicates for luminal CD24^{Hi}Sca1⁺ subset. Five-hundred cells were seeded for each replicate. A colony was defined as a cell cluster of >5 cells. Bar graphs show means \pm s.e.m.; two-sided unpaired Student's t -test; * $P < 0.05$. **g**, Bar graphs showing number of spheres derived from FVB-control and *PIK3CA*^{H1047R}-expressing luminal (CD24^{Hi}Sca1^{-/+}) mammary cells over three passages. Representative data (three replicates, $n = 4$ mice per genotype) from two independent experiments. Bar graphs show means \pm s.d. * $P < 0.02$, two-sided unpaired Student's t -test. **h**, Representative images of spheres derived from CD24^{Hi}Sca1⁻ cells in passage one (P1) and three (P3). Scale bars, 100 μ m. N.d., not determined; NS, not significant.



Extended Data Figure 9 | *PIK3CA*^{H1047R}-evoked tumours express basal and luminal markers. **a**, Representative FACS plots of Lgr5-CreER^{T2}/*PIK3CA*^{H1047R} and K8-CreER^{T2}/*PIK3CA*^{H1047R} tumours ($n = 3$). **b**, Percentages of total GFP-positive cells and GFP-positive basal (CD24^{Lo}Sca1⁺) and luminal (CD24^{Hi}Sca1⁺) subsets of Lgr5-CreER^{T2} and K8-CreER^{T2}/*PIK3CA*^{H1047R} tumours ($n = 3$). Bar graphs show means \pm s.e.m. NS, not significant; two-sided unpaired Student's *t*-test **c**, Immunostaining for basal and luminal markers on serial sections of a multi-nodular rosette-type adenomyoepithelioma (Lgr5-CreER^{T2}/*PIK3CA*^{H1047R}) and adenomyoepithelioma (K8-CreER^{T2}/*PIK3CA*^{H1047R}). Scale bars, 100 μ m, 50 μ m

(magnifications). **d**, Quantification of basal- and luminal-lineage markers of Lgr5-CreER^{T2} and K8-CreER^{T2}/*PIK3CA*^{H1047R} tumours. Each dot represents one tumour (top: K8/18, K14 and SMA $n = 15$, K5 $n = 14$, ER $n = 10$, PR $n = 9$, p63 $n = 8$; bottom: K8/18, K14, SMA and K5 $n = 15$, ER, PR and p63 $n = 10$). All Lgr5-CreER^{T2}/*PIK3CA*^{H1047R} tumours and 8/10 and 6/10 of K8-CreER^{T2}/*PIK3CA*^{H1047R} tumours show more than 1% of ER- and/or PR-positive cells, respectively. Bar graphs show means \pm s.d. **e**, Representative haematoxylin and eosin stainings of tumour phenotypes. Scale bars, 100 μ m. **f**, Percentage of benign and malignant mammary tumours.



Extended Data Figure 10 | Expression profiling of K8- and Lgr5-CreER^{T2}/*PIK3CA*^{H1047R} mammary tumours. **a**, Principle component analysis and dendrogram of a hierarchical clustering of gene expression profiles from 10 K8- and 10 Lgr5-CreER^{T2}/*PIK3CA*^{H1047R} tumours and 2–3 reference mammary glands. Each dot indicates one sample. Circles represent

K8-CreER^{T2} and squares represent Lgr5-CreER^{T2} animals expressing *PIK3CA*^{H1047R} (filled symbols) or not (open symbols). **b**, Heat map of the top 1,000 genes that vary between K8-CreER^{T2}/*PIK3CA*^{H1047R} and Lgr5-CreER^{T2}/*PIK3CA*^{H1047R} tumours and The Cancer Genome Atlas (TCGA) human breast cancer gene signatures. Lum, luminal.

Reactivation of multipotency by oncogenic PIK3CA induces breast tumour heterogeneity

Alexandra Van Keymeulen^{1*}, May Yin Lee^{1*}, Marielle Ousset¹, Sylvain Brohée², Sandrine Rorive^{3,4}, Rajshekhar R. Giraddi¹, Aline Wuidart¹, Gaëlle Bouvencourt¹, Christine Dubois¹, Isabelle Salmon^{3,4}, Christos Sotiriou², Wayne A. Phillips^{5,6} & Cédric Blanpain^{1,7}

Breast cancer is the most frequent cancer in women and consists of heterogeneous types of tumours that are classified into different histological and molecular subtypes^{1,2}. *PIK3CA* and *P53* (also known as *TP53*) are the two most frequently mutated genes and are associated with different types of human breast cancers³. The cellular origin and the mechanisms leading to *PIK3CA*-induced tumour heterogeneity remain unknown. Here we used a genetic approach in mice to define the cellular origin of *Pik3ca*-derived tumours and the impact of mutations in this gene on tumour heterogeneity. Surprisingly, oncogenic *Pik3ca*^{H1047R} mutant expression at physiological levels⁴ in basal cells using keratin (K)5-CreER^{T2} mice induced the formation of luminal oestrogen receptor (ER)-positive/progesterone receptor (PR)-positive tumours, while its expression in luminal cells using K8-CreER^{T2} mice gave rise to luminal ER⁺PR⁺ tumours or basal-like ER⁻PR⁻ tumours. Concomitant deletion of *p53* and expression of *Pik3ca*^{H1047R} accelerated tumour development and induced more aggressive mammary tumours. Interestingly, expression of *Pik3ca*^{H1047R} in unipotent basal cells gave rise to luminal-like cells, while its expression in unipotent luminal cells gave rise to basal-like cells before progressing into invasive tumours. Transcriptional profiling of cells that underwent cell fate transition upon *Pik3ca*^{H1047R} expression in unipotent progenitors demonstrated a profound oncogene-induced reprogramming of these newly formed cells and identified gene signatures characteristic of the different cell fate switches that occur upon *Pik3ca*^{H1047R} expression in basal and luminal cells, which correlated with the cell of origin, tumour type and different clinical outcomes. Altogether our study identifies the cellular origin of *Pik3ca*-induced tumours and reveals that oncogenic *Pik3ca*^{H1047R} activates a multipotent genetic program in normally lineage-restricted populations at the early stage of tumour initiation, setting the stage for future intratumoural heterogeneity. These results have important implications for our understanding of the mechanisms controlling tumour heterogeneity and the development of new strategies to block *PIK3CA* breast cancer initiation.

Breast cancers can be classified into different histological and molecular subtypes including luminal (ER⁺ and/or PR⁺), HER2⁺ and basal-like/triple-negative (ER⁻PR⁻HER2⁻) cancers, which are usually associated with different gene expression and mutation profiles, prognosis and response to therapies⁵. *PIK3CA* mutations are found in about 30% of breast cancers, more frequently in luminal tumours, although they are also found in basal-like/triple-negative breast cancers^{3,6–9}. Expression of oncogenic *Pik3ca*^{H1047R} in all mammary gland lineages using MMTV-Cre mice^{4,10–12} or preferentially in luminal progenitors using WAP-Cre mice¹³ induces heterogeneous mammary tumours^{4,10–13}. The reason for this tumour heterogeneity

upon expression of the *Pik3ca*^{H1047R} mutant in the mammary gland is currently unknown.

To determine whether breast tumour heterogeneity is determined by the cancer cell of origin, we developed a genetic strategy allowing the expression of the oncogenic *Pik3ca* mutant at physiological levels using Cre-inducible *Pik3ca*^{H1047R} knock-in mice⁴, specifically in basal cells (BCs) using K5-CreER^{T2} or in luminal cells (LCs) using K8-CreER^{T2} mice¹⁴ and followed their fate and tumorigenic potential over time. Tamoxifen (TAM) was administered at a dose that does not impair long-term mammary gland development and homeostasis, and resulted in the specific labelling of about 20% of BCs (Extended Data Fig. 1) in 4–5-week-old K5-CreER^{T2}/*Pik3ca*^{H1047R} mice (Fig. 1a). While it has been suggested that the mammary gland contains bipotent basal stem cells^{15,16}, our data using K5-CreER^{T2} knock-in or K14-rtTA/TetO-Cre mice, despite the labelling of 20–50% of BCs, showed no contribution of BCs to the luminal lineage (Extended Data Fig. 1). Further lineage-tracing studies that label all BCs or all LCs will be required to determine whether the discrepancy between the different

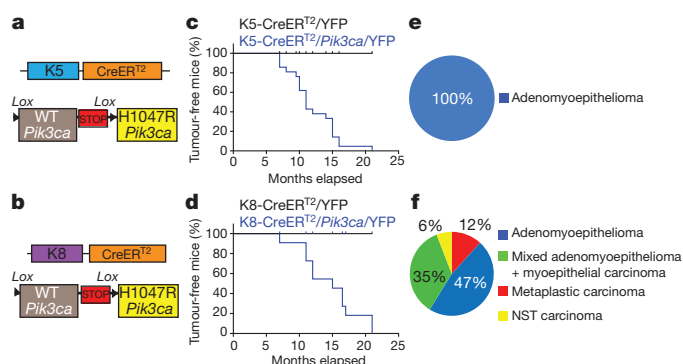


Figure 1 | Oncogenic *Pik3ca* expression in BCs or LCs leads to distinct tumour phenotypes. **a, b**, Genetic strategy to target *Pik3ca*^{H1047R} in BCs (**a**) or LCs (**b**). WT, wild type. **c, d**, Tumour-free survival curves in K5-CreER^{T2}/*Pik3ca*^{H1047R}/Rosa26-YFP mice ($n = 24$ mice) (latency of 12 ± 4 months (mean \pm s.d.)) (**c**) or K8-CreER^{T2}/*Pik3ca*^{H1047R}/Rosa26-YFP mice ($n = 11$ mice) (latency of 15 ± 4 months) (**d**). No tumours were observed in control K5-CreER^{T2}/Rosa26-YFP mice ($n = 10$ mice) (**c**) or in K8-CreER^{T2}/Rosa26-YFP mice ($n = 10$ mice) (**d**). **e, f**, Pie chart showing tumour classification in BC- and LC-derived (**f**) tumours. BC-derived tumours were all classified as adenomyoepitheliomas ($n = 36$ tumours). LC-derived tumours comprised adenomyoepitheliomas ($n = 8$ tumours), mixed adenomyoepithelioma with myoepithelial carcinoma ($n = 6$ tumours), metaplastic carcinoma ($n = 2$ tumours) and invasive carcinoma of NST ($n = 1$ tumour) (**f**). Detailed histological characterization is presented in Extended Data Fig. 2.

¹Université Libre de Bruxelles, IIRBHM, Brussels B-1070, Belgium. ²Institut Jules Bordet, Université Libre de Bruxelles, Brussels B-1000, Belgium. ³Department of Pathology, Erasme Hospital, Université Libre de Bruxelles, Brussels B-1070, Belgium. ⁴DIAPATH—Center for Microscopy and Molecular Imaging (CMMI), Gosselies B-6041, Belgium. ⁵Surgical Oncology Research Laboratory, Peter MacCallum Cancer Centre, Melbourne 3002, Australia. ⁶Sir Peter MacCallum Department of Oncology, University of Melbourne, Parkville 3002, Australia. ⁷WELBIO, Université Libre de Bruxelles, Brussels B-1070, Belgium.

*These authors contributed equally to this work.

studies arises from the unspecific and simultaneous labelling of BCs and LCs. BC-derived mammary tumours arose with a latency of about 12 ± 4 months (mean \pm standard deviation (s.d.)) and were all luminal-like tumour cells that were ER⁺PR⁺, surrounded by BCs (Fig. 1c, e and Extended Data Fig. 2a–e), classified by pathologists as adenomyoepithelioma¹⁷ in mouse and humans¹⁷ (Extended Data Fig. 3a–d). Principal component analysis (PCA) and gene clustering analysis of gene expression profile from fluorescence-activated cell sorting (FACS)-isolated tumour cells using the PAM50 gene set showed that these BC-derived tumours clustered together with the luminal B breast cancer subtype (Extended Data Figs 3 and 4).

The same dose of TAM was administered to 4–5-week-old K8-CreER^{T2}/Pik3ca^{H1047R} mice, resulting in the specific labelling of about 20–30% of LCs (Fig. 1b and Extended Data Fig. 1). Mammary tumours arose with a similar latency (15 ± 4 months) (Fig. 1d). Histological and immunofluorescence analysis revealed that these tumours were more heterogeneous, more aggressive and more proliferative than BC-derived tumours. These tumours comprised adenomyoepithelioma, mixed adenomyoepithelioma with myoepithelial carcinoma, invasive carcinoma of no special type (NST), as well as tumours that show features of metaplastic basal-like breast cancers similar to human breast cancers (Fig. 1f and Extended Data Figs 2, 3). Principal component and gene expression clustering analyses from cells isolated from seven different luminal-derived tumours showed that ER⁺ tumours clustered together with luminal human breast cancers, NST tumours clustered in between luminal B and HER2⁺ tumours, and metaplastic carcinoma clustered with basal-like or HER2⁺ cancers depending on the clustering algorithm (Extended Data Fig. 3j, l), consistent with the phenotypic heterogeneity of the tumours. These results revealed that Pik3ca^{H1047R} expression in LCs gives rise to distinct types of tumours that are generally more aggressive compared with BC-derived tumours. The greater tumour heterogeneity found in the LC-derived tumours may arise from the greater plasticity of LCs and/or the heterogeneity of the luminal progenitor populations initially targeted in the K8-CreER^{T2} mice.

We then assessed whether concomitant p53 deletion affects the phenotype of mammary tumours depending on their cellular origin. K5-CreER^{T2}/Pik3ca^{H1047R}/p53^{fl/fl} mice treated with TAM rapidly developed skin and other cancers that required terminating the experiment before they developed mammary tumours (data not shown). To circumvent this problem, we used mice heterozygous for p53 (K5-CreER^{T2}/Pik3ca^{H1047R}/p53^{fl/+}) and another basal Cre driver (K14-rtTA/TetO-Cre/Pik3ca^{H1047R}/p53^{fl/+}) that alleviated the increased early mortality seen with the K5-CreER^{T2}/Pik3ca^{H1047R}/p53^{fl/fl} mice. BC-derived p53 heterozygous tumours arose with a latency of 9 ± 2 months and consisted mostly of adenomyoepithelioma luminal-like tumours (42–75%), as well as myoepithelial carcinoma (0–16%), NST tumours (0–12%) and metaplastic carcinoma (12–42%) (Fig. 2 and Extended Data Fig. 5). As previously shown using MMTV-Cre mice¹⁰, Pik3ca^{H1047R} expression together with p53 deletion in LCs dramatically accelerates tumour formation, with a latency of 5 ± 1 months for p53 homozygous and 9 ± 3 months for p53 heterozygous mice (Fig. 2). In contrast to BCs, LC-derived p53-deficient tumours always consisted of aggressive carcinomas consisting mostly of metaplastic carcinoma and high-grade myoepithelial carcinoma with characteristics of epithelial-to-mesenchymal transition (Fig. 2 and Extended Data Fig. 5), as previously reported following Pik3ca^{H1047R} expression in all mammary gland cells^{10,12} and found in human basal-like breast cancers with activation of the PI3K pathway by somatic PIK3CA mutations and gene copy number amplification^{3,7–9}. Gene expression clustering of these tumours using the PAM50 genes showed that they clustered together with human basal-like or HER2⁺ subtypes depending on the clustering algorithm (Extended Data Fig. 3k, l). These data demonstrate that concomitant Pik3ca^{H1047R} expression and p53 deletion accelerates tumour development in basal and luminal lineages

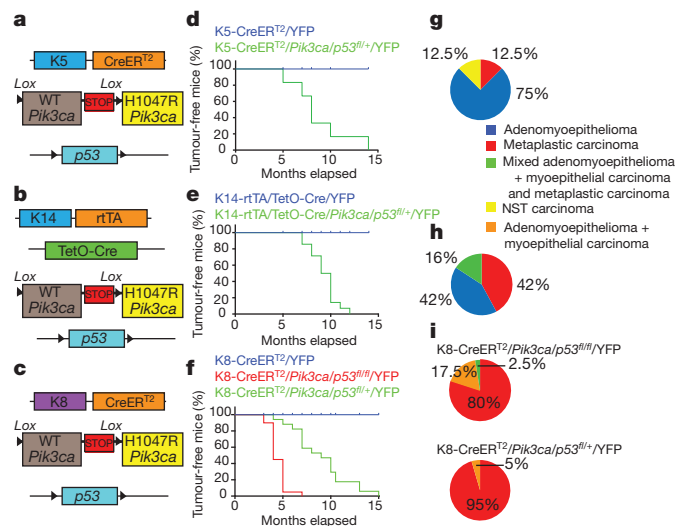


Figure 2 | Oncogenic *Pik3ca* expression and p53 deletion in BCs or LCs leads more frequently to highly invasive mammary tumours. **a–c**, Genetic strategy to target *Pik3ca*^{H1047R} expression and p53 deletion in BCs (**a**, **b**) or in LCs (**c**). **d–f**, Tumour-free survival curves in K5-CreER^{T2}/Pik3ca^{H1047R}/p53^{fl/+}/Rosa26-YFP (9 ± 3 months latency; mean \pm s.d.) ($n = 6$ mice) (**d**), K14-rtTA/TetO-Cre/Pik3ca^{H1047R}/p53^{fl/+}/Rosa26-YFP (9 ± 1 month latency) ($n = 14$ mice) (**e**), K8-CreER^{T2}/Pik3ca^{H1047R}/p53^{fl/+}/Rosa26-YFP mice (5 ± 1 month latency) ($n = 20$ mice) and K8-CreER^{T2}/Pik3ca^{H1047R}/p53^{fl/+}/Rosa26-YFP mice (9 ± 3 months latency) ($n = 17$ mice) (**f**). Control mice did not develop tumours ($n = 10$ mice per condition). **g–i**, Pie charts depicting the classification of mammary tumours in K5-CreER^{T2}/Pik3ca^{H1047R}/p53^{fl/+}/Rosa26-YFP ($n = 8$ tumours) (**g**), K14-rtTA/TetO-Cre/Pik3ca^{H1047R}/p53^{fl/+}/Rosa26-YFP ($n = 19$ tumours) (**h**), K8-CreER^{T2}/Pik3ca^{H1047R}/p53^{fl/+}/Rosa26-YFP ($n = 40$ tumours) and K8-CreER^{T2}/Pik3ca^{H1047R}/p53^{fl/+}/Rosa26-YFP ($n = 22$ tumours) (**i**). Detailed histological characterization is presented in Extended Data Fig. 5.

and that very aggressive metaplastic tumours arise more frequently from oncogenic targeting of LCs than from BCs.

To define further the cellular basis of intratumoural heterogeneity found in *Pik3ca*^{H1047R}-derived tumours, we combined Rosa26-YFP lineage tracing and *Pik3ca*^{H1047R} expression specifically in LCs or BCs and assessed cell fate change over time. Interestingly, as early as 5 weeks after *Pik3ca*^{H1047R} expression in LCs, yellow fluorescent protein (YFP) was also detected in basal-like cells clustered around LCs (Fig. 3a–e and Extended Data Fig. 6), while, as previously described¹⁴, K8-CreER^{T2}-targeted cells consist of a self-sustained unipotent population of LCs (Extended Data Fig. 6a–e). Clonal analysis of LCs expressing oncogenic PIK3CA revealed the presence of bipotent clones containing adjacent LCs and BCs, which were never observed in YFP control LCs (Fig. 3f and Extended Data Fig. 6n–p). The relatively small proportion of K8⁺/K5⁺ BCs compared with K8[−]/K5⁺ BCs suggests that in the initial stage of LC-to-BC transition, these cells expressed markers of both lineages before maturing into basal-like cells and losing expression of LC markers (Fig. 3g and Extended data Fig. 6q–t), which is consistent with the sequential gene expression shown by quantitative polymerase chain reaction with reverse transcription (qRT-PCR) analysis of FACS-isolated BCs and LCs after *Pik3ca*^{H1047R} expression (Extended Data Fig. 6u, v). The proportion of YFP-expressing LCs increased over time, as well as the proportion of YFP⁺ BCs (Fig. 3e), suggesting that *Pik3ca*^{H1047R} confers a competitive advantage on luminal targeted cells. To determine functionally whether LCs acquired multipotency upon PIK3CA expression, we tested the ability of *Pik3ca*^{H1047R}-expressing LCs and their BC progeny to reconstitute the mammary gland upon transplantation into mammary fat pads. FACS-isolated LCs expressing *Pik3ca*^{H1047R} were able to form outgrowths of mammary epithelium containing both BCs and LCs (observed in 6 out of 28 transplants), while, as previously

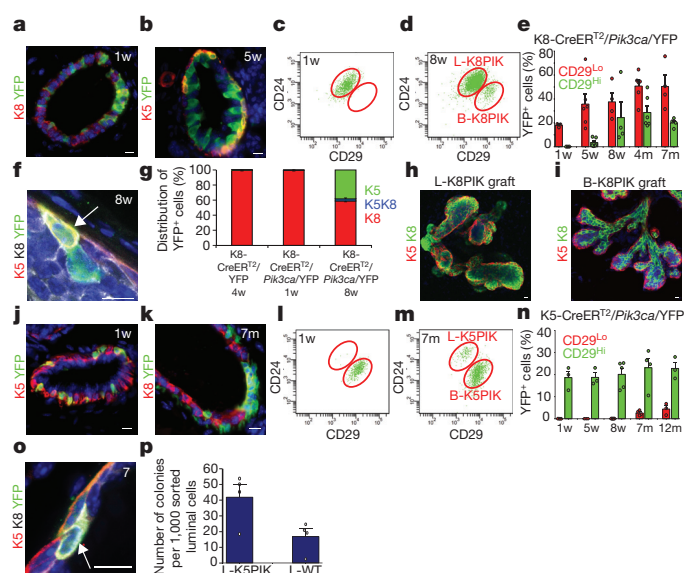


Figure 3 | Oncogenic *Pik3ca* expression induces multipotency in unipotent luminal and basal progenitors. **a, b**, Immunofluorescence of K8/YFP at 1 week (**a**) and K5/YFP at 5 weeks (**b**) after TAM administration to K8-CreER^{T2}/*Pik3ca*^{H1047R}/Rosa26-YFP mice. **m**, month; **w**, week. **c, d**, FACS analysis of CD24 and CD29 expression in Lin-YFP⁺ cells 1 week (**c**) or 8 weeks (**d**) after TAM induction. **e**, Percentage of YFP⁺ cells within LCs (CD29^{Lo}/CD24⁺) and BCs (CD29^{Hi}/CD24⁺) at different time points after TAM administration ($n = 3, 6, 3, 6, 4$ mice for 1 week, 5 weeks, 8 weeks, 4 months and 7 months, respectively). **f**, Immunofluorescence of K8/K5/YFP 8 weeks after clonal induction of K8-CreER^{T2}/*Pik3ca*^{H1047R}/Rosa26-YFP mice. Arrow points to K5/K8/*Pik3ca*^{H1047R}/YFP⁺ BC newly generated from a LC. **g**, Percentage of YFP⁺ cells expressing K5 and/or K8 at different time points after *Pik3ca*^{H1047R} expression in LCs ($n = 3$ mice per condition). See Methods for more details. **h, i**, Immunofluorescence of K5/K8 of a mammary outgrowth derived from LCs (**h**) or BCs (**i**) from K8-CreER^{T2}/*Pik3ca*^{H1047R}/Rosa26-YFP mice. **j, k**, Immunofluorescence of K5/YFP 1 week (**j**) or of K8/YFP 7 months (**k**) after TAM administration to K5-CreER^{T2}/*Pik3ca*^{H1047R}/Rosa26-YFP mice. **l, m**, FACS analysis of CD24 and CD29 expression in Lin-YFP⁺ cells 1 week (**l**) or 7 months (**m**) after TAM induction. **n**, Percentage of YFP⁺ cells within LCs and BCs at different time points after TAM administration ($n = 3, 3, 5, 4, 4, 3$ mice for 1 week, 5 weeks, 8 weeks, 7 months and 12 months, respectively). **o**, Immunofluorescence for K8/K5/YFP 7 months after clonal *Pik3ca*^{H1047R}/YFP expression in BCs. Arrow points to newly formed K8⁺YFP⁺ LC arising from a BC. **p**, Mean number of colonies per 1,000 sorted luminal cells in an *in vitro* colony-forming assay of YFP⁺ LCs derived from K5-CreER^{T2}/*Pik3ca*^{H1047R}/Rosa26-YFP mice induced for 12 months or wild-type LCs ($n = 3$ biologically independent experiments per condition). Circles, individual data points. Error bars, standard error of the mean (s.e.m.). Scale bars, 10 μ m.

described^{14,18,19}, wild-type LCs derived from K8-CreER^{T2}/Rosa26-YFP mice were not able to form mammary outgrowths in the same conditions (in 0 out of 10 transplants) (Fig. 3h). Likewise, transplantation of newly formed BCs from K8-CreER^{T2}/*Pik3ca*^{H1047R} mice also generated mammary outgrowths containing BCs and LCs (7/11), as efficiently as control BCs (8/10) (Fig. 3i). Altogether these data show that oncogenic *Pik3ca* promotes multilineage differentiation of LCs, inducing cellular heterogeneity at the early stage of the tumour initiation process.

In contrast to the early multilineage differentiation observed after oncogenic PIK3CA expression in LCs, during the first few months after TAM administration to K5-CreER^{T2}/*Pik3ca*^{H1047R}/Rosa26-YFP mice, BCs remained unipotent. Only around 7 months after oncogene expression, newly formed LCs became detectable and progressively increased over time (Fig. 3j–o and Extended Data Fig. 7). Immunostaining and qRT-PCR showed that these newly formed LCs expressed luminal markers at similar levels to wild-type LCs

and no longer expressed high levels of basal markers (Extended Data Fig. 7k, l). To determine whether LCs derived from BCs functionally correspond to luminal progenitors, we assessed the clonogenic potential of these cells using a colony-forming assay that only allows the growth of LCs²⁰. The number of colonies derived from FACS-isolated LCs after *Pik3ca*^{H1047R} expression in BCs was significantly higher compared with wild-type LCs, supporting the notion that oncogenic PIK3CA promotes the reprogramming of BCs into functional LCs (Fig. 3p). Altogether, these data show that *Pik3ca*^{H1047R} induces multipotency in otherwise lineage-restricted basal and luminal unipotent progenitors, inducing cellular heterogeneity in oncogene-targeted cells before progressing into more invasive tumours.

To define the molecular mechanisms by which *Pik3ca*^{H1047R} promotes multipotency and tumour heterogeneity, we performed transcriptional profiling of FACS-isolated basal-like and luminal-like cells after *Pik3ca*^{H1047R} induction in LCs (K8-CreER^{T2}/*Pik3ca*^{H1047R}/Rosa26-YFP mice) (B-K8PIK and L-K8PIK) and in BCs (K5-CreER^{T2}/*Pik3ca*^{H1047R}/Rosa26-YFP mice) (B-K5PIK and L-K5PIK) (Fig. 3d, m and Supplementary Table 1). Gene expression clustering analysis showed that a profound reprogramming occurred in PIK3CA-(H1047R)-expressing cells as they underwent transition between basal and luminal lineages, becoming molecularly similar to the mammary lineages that they were converted into (Fig. 4a).

To unravel the molecular mechanisms by which oncogenic *Pik3ca* induced changes in cell fate and determine whether these mechanisms are conserved or distinct across different cells of origin, we defined the gene signature induced by the expression of oncogenic *Pik3ca* in each population (that is, B-K5PIK versus B-K5YFP, L-K5PIK versus L-K8YFP, and so on). Only three genes were upregulated in all conditions (*Serpina3n*, *Gdpd3* and *Zfp949*), and most of the genes upregulated by *Pik3ca*^{H1047R} expression were dependent on both the origin of the cell in which *Pik3ca*^{H1047R} was initially expressed and the cell lineage in which the oncogene was currently expressed (Fig. 4b, Supplementary Table 2 and Extended Data Fig. 8). While *Pik3ca*^{H1047R} induced the expression of specific genes according to their cellular origin and their basal or luminal phenotypes, 51 annotated genes were commonly upregulated in L-K5PIK and in B-K8PIK, including the long non-coding RNA *Neat1* and the transcription factor *Runx2*, which both regulate mammary cell fate^{21,22}, genes regulating signal transduction (for example, *Sfrp2*), cellular metabolism (for example, *Tktl1*, *Bdh1*) and cell adhesion (for example, *EphB2*, *Trio*) (Fig. 4b, c and Supplementary Table 2), suggesting that common and distinct mechanisms induce cell fate changes upon oncogenic *Pik3ca* expression.

BC-to-LC fate transition induced by oncogenic *Pik3ca* induced the expression of a distinct set of genes (basal-to-luminal multipotent signature) (Fig. 4). Some of these genes, such as *Ntrk3* (also known as *TrkC*), were already upregulated in BCs after *Pik3ca*^{H1047R} expression (Supplementary Table 2), suggesting that they represent the signature of their BC of origin. After BC to LC fate transition, *Pik3ca*^{H1047R} expression induced genes that are specific for the newly formed LCs, including *Nrtk2* (also known as *TrkB*), a neurotrophin receptor expressed in wild-type BCs and transiently expressed in LCs during *Pik3ca*^{H1047R}-induced BC-to-LC fate transition (Fig. 4d and Extended Data Fig. 7m). NRTK2 and NRTK3 have previously been shown to be expressed in breast cancers and regulate survival of breast cancer stem cells in response to chemotherapy^{23,24}. In addition, a translocation leading to a gene fusion between *Etv6* and *Ntrk3* causes breast tumours in both mouse and humans^{25,26}. *Etv6-Ntrk3* expression in the mammary gland predominantly induces the same type of bipotent mammary tumours²⁶ that arise from *Pik3ca*^{H1047R} expression in BCs, supporting a role for the *Pik3ca/Ntrk2-Ntrk3* axis in the establishment of bipotency during breast tumour initiation. The basal-to-luminal upregulated multipotent signature also contained genes commonly upregulated in L-K8PIK (Extended Data Fig. 8k), reflecting the consequence of PIK3CA expression in LCs.

In contrast, the luminal-to-basal multipotent signature was characteristic of a wounding and proliferation response marked by the upregulation of *Il33* (also known as alarmin; a cytokine that has been shown to be overexpressed in breast cancers and attenuates NK response against tumour cells^{27,28}), *Il24a*, *Krt16*, *Itgb6*, *Itga2*, *Itga5*, *Tnc*, *Cd109*, *Plau*, *Wnt10a*, *Timp3*, *Inhba*, *Ngf*, *Ereg*, *Ccnd1* and *Ccnd2* (Fig. 4e). As found during BC-to-LC transition, most of the luminal-to-basal multipotent signature genes were specific for the newly formed BCs (for example, *Ereg*, *Ccnd1*, *Wnt10a*, *Il33*); a significant fraction of these genes (for example, *Krt16*, *Il24a*, *Ccnd2*, *Inhba*, *Tnc*) were already upregulated in LCs targeted by oncogenic PIK3CA, suggesting that they represent the signature of the LC of origin (Supplementary Tables 2, 3 and Extended Data Fig. 8).

To define the relevance of these multipotency gene signatures to tumour progression, we assessed the expression of these genes in *Pik3ca*^{H1047R}-derived tumours. Some luminal-to-basal multipotent signature genes such as *Il24a*, *Krt16* and *Plau* were only upregulated during the initial stage of reprogramming and downregulated thereafter, while other genes such as *Col11a1*, the epidermal growth factor

receptor (EGFR) ligand *Ereg*, *Inhba*, *Wnt10a* and *Tnc* continued to be expressed, or even further increased, in basal-like breast cancers arising from LCs (Fig. 4e, f). Similarly, *Ntrk2* and *Ntrk3* were expressed or even further upregulated in K5-CreER^{T2}/*Pik3ca*^{H1047R}-derived luminal tumours (Fig. 4f). These data indicate that some of the genes associated with cell fate transition during the early steps of tumour initiation increase with tumour progression.

To define the relevance of the *Pik3ca*^{H1047R}-induced multipotent gene signatures in human breast cancers, we assessed whether the different multipotent signatures correlated with a particular molecular breast cancer subtype²⁹ or disease-free survival in a cohort of systemically untreated breast cancer patients³⁰. Interestingly, the luminal-to-basal transition gene signature was strongly associated with basal-like breast cancers (Fig. 4g). Higher expression levels of this gene signature or individual genes such as *NGF*, *INHBA*, *ITGB6* and *WNT10A* were associated with poor clinical outcome (Fig. 4h and Extended data Fig. 9), consistent with the more aggressive tumour types induced by *Pik3ca*^{H1047R} expression in LCs. In contrast, the BC-to-LC fate signature was associated with luminal A and normal-like human breast cancers (Fig. 4i). High gene expression levels of this gene signature were significantly associated with better prognosis (Fig. 4j), consistent with the less aggressive tumours arising from BCs. These data indicate that the genetic program associated with *Pik3ca*^{H1047R}-induced multipotency correlated with distinct molecular subtypes of human breast cancers and their levels of expression correlated with distinct clinical outcome.

Our study shows that the cell of origin controls tumour heterogeneity in *Pik3ca*^{H1047R}-induced mammary tumours. *Pik3ca*^{H1047R} expression in LCs gives rise to aggressive basal-like tumours while expression in BCs gives rise to less aggressive luminal-like tumours. We demonstrate that *Pik3ca*^{H1047R} induced multipotency in unipotent progenitors. The promotion of multipotency induced by *Pik3ca*^{H1047R} is regulated by common and cell-lineage-specific molecular mechanisms that are influenced by the cellular origin in which the oncogene is initially expressed, setting the stage for future tumour heterogeneity and influencing clinical outcome in patients with breast cancers.

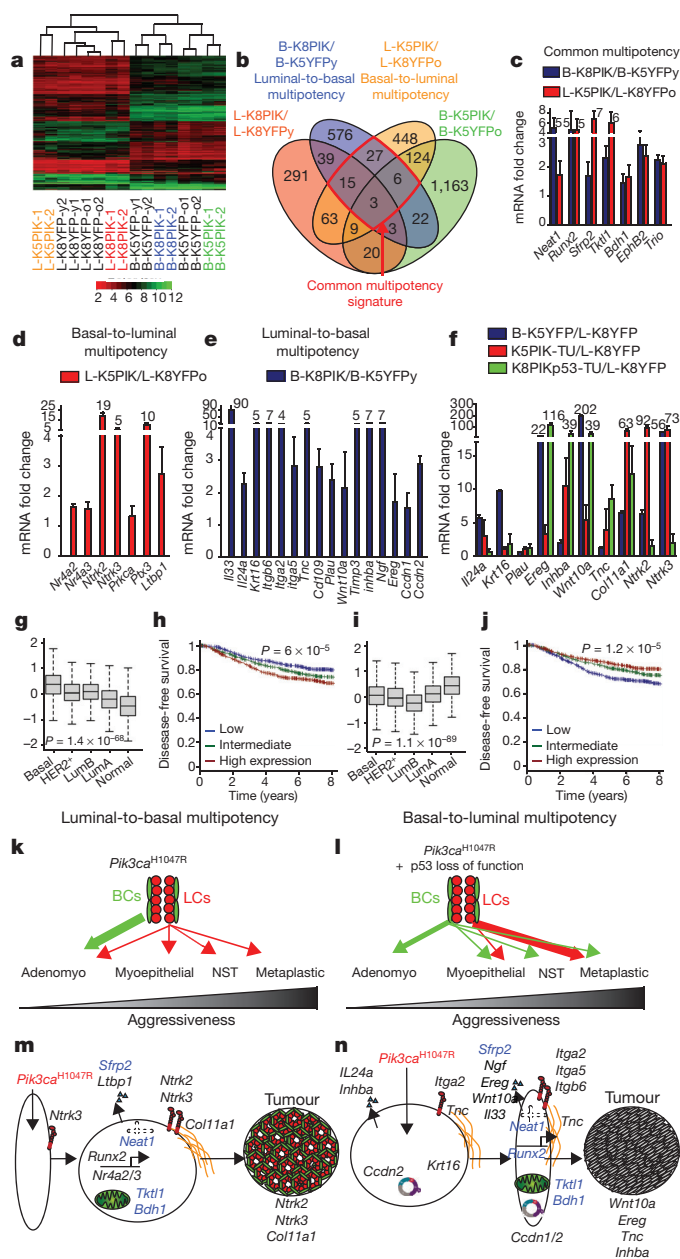


Figure 4 | Molecular characterization of oncogenic *Pik3ca*-induced multipotency. **a**, Hierarchical gene expression clustering of BCs and LCs with or without *Pik3ca*^{H1047R} expression. Green and red correspond to high and low expressed genes, respectively. The two major branches of the tree are supported by bootstrap values of 100. **b**, Venn diagram of upregulated genes (>1.5 fold) after *Pik3ca*^{H1047R} expression in BCs and LCs. **c–e**, qRT-PCR analysis of genes belonging to the common (**c**), basal-to-luminal (**d**), or luminal-to-basal multipotency signature (**e**) in B-K8PIK and L-K5PIK cell population, 8 weeks and 10–12 months after *Pik3ca*^{H1047R} expression, respectively, compared with their age-matched controls. Gene expression was normalized to *Gapdh* housekeeping gene ($n = 4$ biologically independent samples). **f**, qRT-PCR analysis of the multipotency signature genes in control cells, in BC-derived adenomyoepithelioma and in LC-derived metaplastic tumours. Data were normalized to gene expression in age-matched control LCs (L-K8YFPo) ($n = 4$ biologically independent samples). **g**, **i**, Expression levels of the luminal-to-basal (**g**) or basal-to-luminal (**i**) multipotency signature in a large set of breast cancer patients according to their PAM50 subtype. Lum, luminal. **h**, **j**, Disease-free survival in untreated patients according to the level of expression of the genes of the luminal-to-basal (**h**) or basal-to-luminal (**j**) multipotency signature. **k–n**, Summary of the role of the cancer cell of origin in regulating *Pik3ca*^{H1047R}-induced tumour heterogeneity. **k**, Expression of *Pik3ca*^{H1047R} in BCs gives rise to luminal-like tumours, while in LCs *Pik3ca*^{H1047R} gives rise to more heterogeneous and aggressive tumours. Types of carcinoma are noted at the bottom of the panel. Adenomyo, adenomyoepithelioma. **l**, Additional p53 deletion promotes *Pik3ca*^{H1047R}-induced tumour heterogeneity in BCs and leads to more aggressive metaplastic carcinoma in LCs. **m**, **n**, Model of *Pik3ca*^{H1047R}-induced multipotency in LCs and BCs. Genes shown are upregulated during cell fate change. Genes highlighted in blue belong to the common multipotency signature. Error bars, s.e.m.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 30 July 2014; accepted 15 June 2015.

Published online 12 August 2015.

- Weigelt, B., Geyer, F. C. & Reis-Filho, J. S. Histological types of breast cancer: how special are they? *Mol. Oncol.* **4**, 192–208 (2010).
- Perou, C. M. *et al.* Molecular portraits of human breast tumours. *Nature* **406**, 747–752 (2000).
- Network, C. G. A. Comprehensive molecular portraits of human breast tumours. *Nature* **490**, 61–70 (2012).
- Tikoo, A. *et al.* Physiological levels of *Pik3ca*^{H1047R} mutation in the mouse mammary gland results in ductal hyperplasia and formation of ER α -positive tumors. *PLoS ONE* **7**, e36924 (2012).
- Prat, A., Ellis, M. J. & Perou, C. M. Practical implications of gene-expression-based assays for breast oncologists. *Nature Rev. Clin. Oncol.* **9**, 48–57 (2012).
- Shah, S. P. *et al.* The clonal and mutational evolution spectrum of primary triple-negative breast cancers. *Nature* **486**, 395–399 (2012).
- Stemke-Hale, K. *et al.* An integrative genomic and proteomic analysis of PIK3CA, PTEN, and AKT mutations in breast cancer. *Cancer Res.* **68**, 6084–6091 (2008).
- López-Knowles, E. *et al.* PI3K pathway activation in breast cancer is associated with the basal-like phenotype and cancer-specific mortality. *Int. J. Cancer* **126**, 1121–1131 (2010).
- Tilch, E. *et al.* Mutations in EGFR, BRAF and RAS are rare in triple-negative and basal-like breast cancers from Caucasian women. *Breast Cancer Res. Treat.* **143**, 385–392 (2014).
- Adams, J. R. *et al.* Cooperation between *Pik3ca* and p53 mutations in mouse mammary tumor formation. *Cancer Res.* **71**, 2706–2717 (2011).
- Liu, P. *et al.* Oncogenic PIK3CA-driven mammary tumors frequently recur via PI3K pathway-dependent and PI3K pathway-independent mechanisms. *Nature Med.* **17**, 1116–1120 (2011).
- Yuan, W. *et al.* Conditional activation of *Pik3ca*^{H1047R} in a knock-in mouse model promotes mammary tumorigenesis and emergence of mutations. *Oncogene* **32**, 318–326 (2013).
- Meyer, D. S. *et al.* Luminal expression of PIK3CA mutant H1047R in the mammary gland induces heterogeneous tumors. *Cancer Res.* **71**, 4344–4351 (2011).
- Van Keymeulen, A. *et al.* Distinct stem cells contribute to mammary gland development and maintenance. *Nature* **479**, 189–193 (2011).
- Rios, A. C., Fu, N. Y., Lindeman, G. J. & Visvader, J. E. *In situ* identification of bipotent stem cells in the mammary gland. *Nature* **506**, 322–327 (2014).
- Wang, D. *et al.* Identification of multipotent mammary stem cells by protein C receptor expression. *Nature* **517**, 81–84 (2015).
- Hayes, M. M. Adenomyoepithelioma of the breast: a review stressing its propensity for malignant transformation. *J. Clin. Pathol.* **64**, 477–484 (2011).
- Shackleton, M. *et al.* Generation of a functional mammary gland from a single stem cell. *Nature* **439**, 84–88 (2006).
- Stingl, J. *et al.* Purification and unique properties of mammary epithelial stem cells. *Nature* **439**, 993–997 (2006).
- Shehata, M. *et al.* Phenotypic and functional characterisation of the luminal cell hierarchy of the mammary gland. *Breast Cancer Res.* **14**, R134 (2012).
- Standaert, L. *et al.* The long noncoding RNA Neat1 is required for mammary gland development and lactation. *RNA* **20**, 1844–1849 (2014).
- Owens, T. W. *et al.* Runx2 is a novel regulator of mammary epithelial cell fate in development and breast cancer. *Cancer Res.* **74**, 5277–5286 (2014).
- Jin, W. *et al.* TrkC plays an essential role in breast tumor growth and metastasis. *Carcinogenesis* **31**, 1939–1947 (2010).
- Yin, B. *et al.* The TrkB⁺ cancer stem cells contribute to post-chemotherapy recurrence of triple-negative breast cancers in an orthotopic mouse model. *Oncogene* **34**, 761–770 (2015).
- Euhus, D. M., Timmons, C. F. & Tomlinson, G. E. *ETV6-NTRK3*—Trk-ing the primary event in human secretory breast cancer. *Cancer Cell* **2**, 347–348 (2002).
- Li, Z. *et al.* *ETV6-NTRK3* fusion oncogene initiates breast cancer from committed mammary progenitors via activation of AP1 complex. *Cancer Cell* **12**, 542–558 (2007).
- Milovanovic, M. *et al.* IL-33/ST2 axis in inflammation and immunopathology. *Immunol. Res.* **52**, 89–99 (2012).
- Wu, J. *et al.* Identification and functional analysis of 9p24 amplified genes in human breast cancer. *Oncogene* **31**, 333–341 (2012).
- Parker, J. S. *et al.* Supervised risk predictor of breast cancer based on intrinsic subtypes. *J. Clin. Oncol.* **27**, 1160–1167 (2009).
- Haibe-Kains, B. *et al.* A three-gene model to robustly identify breast cancer molecular subtypes. *J. Natl. Cancer Inst.* **104**, 311–325 (2012).

Supplementary Information is available in the online version of the paper.

Acknowledgements C.B. is an investigator of WELBIO, A.V.K. is Chercheur Qualifié of the FNRS, C.S. is Maître de Recherche of the FNRS, M.Y.L. is supported by the Agency for Science, Technology and Research (A*STAR, Singapore) fellowship, M.O. and A.W. are supported by FNRS fellowships, R.R.G. is supported by a TELEVIE fellowship and S.B. is supported by the foundation “Amis de l’institut Jules Bordet”. The Center for Microscopy and Molecular Imaging is supported by the European Regional Development Fund and Wallonia. W.A.P. is supported by project grants from the National Health and Medical Research Council of Australia. This work was supported by the FNRS, TELEVIE, a research grant from the Fondation Contre le Cancer, the ULB fondation, the Fond Yvonne Boël, the Fond Gaston Ithier, the foundation Bettencourt Schueller, the foundation Baillet Latour, and the European Research Council.

Author Contributions C.B. and A.V.K. designed the experiments and performed data analysis. A.V.K., M.Y.L. and M.O. performed all the experiments. S.B. and C.S. performed the bioinformatic analysis of gene expression and comparison with human breast cancer expression and gene amplification on human samples. S.R. and I.S. helped to perform the histological classification of mouse mammary tumours with regard to their similarities with human breast cancers. G.B. provided technical support. C.D. provided technical support for cell sorting. A.W. and R.R.G. helped with some experiments. W.A.P. provided animals and critically reviewed the manuscript. C.B. and A.V.K. wrote the manuscript.

Author Information Microarrays have been deposited in the Gene Expression Omnibus under accession number GSE69290. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to C.B. (Cedric.Blanpain@ulb.ac.be) or A.V.K. (avkeymeu@ulb.ac.be).

METHODS

Mice. Rosa26-YFP mice³¹ were obtained from the Jackson laboratory. K5-CreER^{T2}, and K8-CreER^{T2} mice were described previously¹⁴. K14-rtTA mice³² were provided by E. Fuchs. TetO-Cre mice³³ were provided by A. Nagy. *Pik3ca*^{H1047R} knock-in mice, in which wild-type exon 20 is replaced by H1047R mutant exon 20 upon Cre recombination, were described previously⁴. *p53*^{fl/fl} mice³⁴ were obtained from the National Cancer Institute at Frederick.

All experimental mice used in this study were female, mixed strains and more than 6 weeks old. No statistical methods were used to predetermine sample size. For all experiments presented in this study, the sample size was large enough to measure the effect size. The experiments were not randomized. The investigators were not blinded to allocation during experiments and outcome assessment. Mice designated within the tumour cohort were killed when a palpable mass of maximum 1 cm³ was detected. Mouse colonies were maintained in a certified animal facility in accordance with European guidelines. Ethical protocol was approved by the local ethical committee for animal welfare (CEBEA) from the Université Libre de Bruxelles (protocols 363 and 527).

Targeting expression of YFP and/or PIK3CA(H1047R) and deletion of p53. Four- to five-week-old K5-CreER^{T2}/*Pik3ca*^{H1047R}/Rosa26-YFP, K8-CreER^{T2}/*Pik3ca*^{H1047R}/Rosa26-YFP, K8-CreER^{T2}/*Pik3ca*^{H1047R}/*p53*^{fl/fl}/Rosa26-YFP, K8-CreER^{T2}/*Pik3ca*^{H1047R}/*p53*^{fl/+}/Rosa26-YFP, K5-CreER^{T2}/*Pik3ca*^{H1047R}/*p53*^{fl/+}/Rosa26-YFP, K5-CreER^{T2}/Rosa26-YFP and K8-CreER^{T2}/Rosa26-YFP mice were induced with 15 mg of tamoxifen (TAM) (Sigma; diluted in sunflower seed oil, Sigma) by intraperitoneal injection (3 injections of 5 mg every 3 days). TAM administration induced a transient delay in mammary gland development during puberty but there was no long-term effect on mammary gland development and homeostasis³⁵. Five-week-old K14-rtTA/TetO-Cre/*Pik3ca*^{H1047R}/*p53*^{fl/+}/Rosa26-YFP and K14-rtTA/TetO-Cre/Rosa26-YFP mice were induced by oral administration of doxycycline food diet (1 g kg⁻¹; BIO-SERV) for 5 days. For clonal analyses, 4- to 5-week-old K8-CreER^{T2}/*Pik3ca*^{H1047R}/Rosa26-YFP and K5-CreER^{T2}/*Pik3ca*^{H1047R}/Rosa26-YFP mice were respectively induced with 0.05 mg or 2 mg TAM by intraperitoneal injection. For induction in adult mice, 8-week-old K8-CreER^{T2}/*Pik3ca*^{H1047R}/Rosa26-YFP mice were induced with 15 mg of TAM by intraperitoneal injection (3 injections of 5 mg every 3 days).

Histology and immunostaining on sections. For immunofluorescence, dissected inguinal mammary glands or tumour samples were pre-fixed for 2 h in 4% paraformaldehyde at room temperature. Tissues were washed three times with PBS for 5 min and incubated overnight in 30% sucrose in PBS at 4 °C. Tissues were embedded in OCT and kept at -80 °C. Sections of 5 µm were cut using a HM560 Microm cryostat (Mikron Instruments).

For immunofluorescence, tissue sections were incubated in blocking buffer (5% horse serum, 1% BSA, 0.2% Triton-X in PBS) for 1 h at room temperature. The different primary antibodies combinations were incubated overnight at 4 °C. Sections were then rinsed three times for 5 min in PBS and incubated with corresponding secondary antibodies diluted at 1:400 in blocking buffer for 1 h at room temperature. The following primary antibodies were used: anti-GFP (rabbit, 1:1,000, A11122, Molecular Probes), anti-GFP (chicken, 1:1,000, ab13970, Abcam), anti-K8 (rat, 1:1000, Troma-I, Developmental Studies Hybridoma Bank, University of Iowa), anti-K14 (rabbit, 1:1,000, PRB-155P-0100, Covance), anti-K14 (chicken, 1:1,000, SIG-3476-0100, Covance), anti-K5 (rabbit, 1:1,000, PRB-160P-0100, Covance), anti-K19 (rat, 1:500, Troma-III, Developmental Studies Hybridoma Bank, University of Iowa), anti-ER (rabbit, 1:300, sc-542, Santa Cruz), anti-PR (rabbit, 1:300, sc-7208, Santa Cruz), anti-Her2 (rabbit, 1:300, 2165, Cell Signaling), anti-Ki67 (rabbit, 1:500, ab15580, abcam), anti-E-cadherin (rat, 1:1,000, 14-3249-82, eBioscience), anti-vimentin (rabbit, 1:400, ab92547, Abcam), anti-Nr1h2 (rabbit, 1:500, sc-12, Santa Cruz), anti-p63 (rabbit, 1:100, Mab306-05, Santa Cruz), anti-SMA-Cy3 (mouse, 1:500, C6198, Sigma-Aldrich), anti-claudin 3 (rabbit, 1:300, 34-1700, Invitrogen). The following secondary antibodies were used: anti-rabbit, anti-rat, anti-chicken conjugated to AlexaFluor488 (Molecular Probes), to Rhodamine Red-X or to Cy5 (JacksonImmunoResearch). Nuclei were stained with Hoechst solution (1:2,000) and slides were mounted in DAKO mounting medium supplemented with 2.5% Dabco (Sigma).

For paraffin-embedded tissues, dissected mammary glands were pre-fixed overnight, 4 °C in paraformaldehyde 4%. Tissues were washed three times with PBS. Prior to automated paraffin processing, tissues were washed in tap water and kept in isopropanol 70%. Five-micrometre sections were made with a Leica RM2245 microtome.

Haematoxylin and eosin staining was performed as previously described³⁶. p63 staining on tumour paraffin sections were performed on an automated IHC platform (Ventana Discovery XT). Briefly, paraffin sections were deparaffinized and rehydrated. The antigen unmasking procedure was performed for 36 min at 95 °C in EDTA (pH 9). Slides were incubated with the anti-p63 (clone 7JUL, 1:100, Leica) for 3 h, followed by a linker rabbit anti-mouse (clone M1gG51-4, abcam

1:750) for 16 min. Finally, slides were incubated with the OmniMap HRP-conjugated anti-rabbit antibody (Ventana) for 12 min. Standard ABC kit, and ImmPACT DAB (Vector Laboratories) were used for the detection of HRP activity. Nuclei staining was done with Mayer's Hematoxylin (Labonord), followed by dehydration and mounting with SafeMount (Labonord).

Whole-mount mammary gland immunofluorescence. For clonal analyses, dissected inguinal mammary glands were incubated in 2 ml HBSS plus 30 U ml⁻¹ collagenase plus 300 µg ml⁻¹ hyaluronidase (Sigma) for 30 min at 37 °C under agitation. After three washes of 5 min with HBSS, mammary glands were fixed in 4% paraformaldehyde for 2 h at room temperature, washed three times for 10 min in PBS under agitation and incubated in blocking buffer (5% horse serum, 1% BSA, 0.8% Triton-X in PBS) for 3 h at room temperature. The primary antibody combination, diluted in the blocking buffer, was incubated overnight at room temperature under agitation. Samples were washed three times for 10 min in PBS/0.2% Tween-20 and incubated in secondary antibodies diluted in the blocking buffer for 5 h under agitation. Cell nuclei were stained with Hoechst for 30 min (1:1,000 in PBS/0.2% Tween-20). Samples were mounted on slides in DAKO mounting medium supplemented with 2.5% Dabco (Sigma).

Staining on human breast cancer sections. Tissue samples were obtained retrospectively from archival formalin-fixed and paraffin-embedded samples in the Department of Pathology of the Erasme Hospital. Histopathological diagnoses were reviewed and assessed according to the 2012 World Health Organization Classification. Sections of 5 µm were subjected to standard immunohistochemistry (IHC) as previously described³⁶ using respectively monoclonal anti-CK8/18 (1:200; clone 5D3; BioGenex), anti-CK14 (1:100; clone LL002; Leica) and anti-P63 (1:200; clone 7JUL; Leica) antibodies. Staining was visualized with streptavidin-biotin-peroxidase complex kit reagents (BioGenex) using diaminobenzidine/H₂O₂ as the chromogenic substrate. Counterstaining with haematoxylin concluded the processing. Nuclei staining was done with Mayer's Haematoxylin (Labonord), followed by dehydration and mounting with SafeMount (Labonord).

Microscope image acquisition. Pictures were acquired on an Axio Observer Z1 Microscope using ×10 and ×40 Zeiss EC Plan-NEOFLUAR objectives, with an AxioCamMR3 camera and using the Axiovision software (Carl Zeiss). Confocal images in Fig. 3f, h, i, o and Extended Data Fig. 6f, h-l, n, o, q-t and Extended Data Fig. 7i, m, were acquired at room temperature using a Zeiss LSM780 multiphoton confocal microscope fitted on an Axiovert M200 inverted microscope equipped with C-Apochromat (×40 = 1.2 numerical aperture) water immersion objectives (Carl Zeiss). Optical sections of 1,024 × 1,024 pixels, were collected sequentially for each fluorochrome. The data sets generated were merged and displayed with the ZEN software.

Mammary gland and tumour cell dissociation. Mammary glands were dissected and lymph nodes removed. Tissues were briefly washed in HBSS, and chopped with a McIlwain tissue chopper. Chopped tissues were placed in HBSS plus 300 U ml⁻¹ collagenase (Sigma) plus 300 µg ml⁻¹ hyaluronidase (Sigma) and digested for 2 h at 37 °C under agitation. Physical dissociation using a P1000 pipette was done every 15 min throughout the enzymatic digestion duration. EDTA at a final concentration of 5 mM was added for 10 min to the resultant organoid suspension, followed by 0.25% Trypsin-EGTA for 2 min (only in the case of normal mammary glands) before filtration through a 70 µm mesh, two successive washes in 2% FBS/PBS and antibody labelling.

Cell labelling, flow cytometry and sorting. Two-to-five million cells per condition were incubated in 250 µl 2% FBS/PBS with fluoro-chrome-conjugated primary antibodies for 30 min, vortexing every 10 min. Cells were washed with 2% FBS/PBS and were resuspended in 2.5 µg ml⁻¹ 4',6-diamidino-2-phenylindole (DAPI; Invitrogen) before analysis. Primary antibodies used were: PE-Cy7-conjugated anti-CD24 (1:50, clone M1/69, BD Biosciences), APC-conjugated anti-CD29 (1:50, clone eBioHmb1-1, eBiosciences), PE-conjugated anti-CD45 (1:50, clone 30-F11, eBiosciences), PE-conjugated anti-CD31 (1:50, clone MEC 13.33, BD Biosciences), PE-conjugated anti-CD140a (1:50, clone APA5, eBiosciences). Data analysis and cell sorting were performed on a FACSaria sorter using the FACS DiVa software (BD Biosciences). Dead cells were excluded with DAPI; CD45-, CD31- and CD140a-positive cells were excluded (Lin⁻) before analysis of the YFP⁺ cells. For profile analysis, a minimum of 1,000 YFP⁺ cells were analysed per sample.

Tumour harvesting and classification. Tumours were detected by mammary gland palpation. Mice were killed when one tumour reached a maximum of 1 cm diameter. The K5-CreER^{T2}/*Pik3ca*^{H1047R}/Rosa26-YFP mice presented 1 tumour in 58%, 2 tumours in 25%, and 3 or more tumours in 17% of the cases at the time of analysis (a total of 36 tumours from 24 mice were analysed). The K8-CreER^{T2}/*Pik3ca*^{H1047R}/Rosa26-YFP mice presented 1 tumour in 64%, 2 tumours in 18%, and 3 or more tumours in 18% of the cases at the time of analysis (a total of 17 tumours from 11 mice were analysed). The K5-CreER^{T2}/*Pik3ca*^{H1047R}/*p53*^{fl/+}/Rosa26-YFP mice presented 1 tumour in 83%, and 3 tumours in 17% of the cases

at the time of analysis (a total of 8 tumours from 6 mice were analysed). The K14-rtTA/TetO-Cre/*Pik3ca*^{H1047R}/*p53*^{fl/+}/Rosa26-YFP mice presented 1 tumour in 57%, 2 tumours in 36%, and 3 or more tumours in 7% of the cases at the time of analysis (a total of 19 tumours from 14 mice were analysed). The K8-CreER^{T2}/*Pik3ca*^{H1047R}/*p53*^{fl/fl}/Rosa26-YFP mice presented 1 tumour in 20%, 2 tumours in 45%, and 3 or more tumours in 35% of the cases at the time of analysis (a total of 40 tumours from 20 mice were analysed). The K8-CreER^{T2}/*Pik3ca*^{H1047R}/*p53*^{fl/+}/Rosa26-YFP mice presented 1 tumour in 71%, and 2 tumours in 29% of the cases at the time of analysis (a total of 22 tumours from 17 mice were analysed). For each harvested tumour, the tumour was cut in three pieces, one for paraffin embedding, one for OCT embedding, and one for cell sorting and RNA extraction. Tumour classification was done based on histological features.

Mammary colony-forming assay. Luminal YFP⁺ cells from K5-CreER^{T2}/*Pik3ca*^{H1047R}/Rosa26-YFP mice induced for 12 months were flow-sorted as a single-cell suspension based on their Lin⁻ CD29^{Lo}CD24⁺YFP⁺ profile. Control YFP⁻ luminal cells from K5-CreER^{T2}/Rosa26-YFP induced for 12 months were sorted based on their CD29^{Lo}CD24⁺ profile. Luminal cells were cultured with irradiated NIH 3T3 feeder cells in Mouse-Epicult B media (Stem Cell Technologies) supplemented with 10 ng ml⁻¹ epidermal growth factor (Sigma-Aldrich), 10 ng ml⁻¹ basic fibroblast growth factor (R&D Systems), 4 µg ml⁻¹ heparin (Sigma-Aldrich), 1 mg ml⁻¹ bovine serum albumin (BSA; Sigma-Aldrich), 5% FBS (Life Technologies), 50 units ml⁻¹ penicillin and 50 µg ml⁻¹ streptomycin (Life Technologies), as previously described²⁰. After 1 week, colonies were fixed with methanol, stained with Giemsa stain (Sigma-Aldrich) and counted manually.

Mammary fat pad transplantation and analysis. Eight thousand LCs from K8-CreER^{T2}/*Pik3ca*^{H1047R}/Rosa26-YFP or control K8-CreER^{T2}/Rosa26-YFP or 1,350 BCs from K8-CreER^{T2}/*Pik3ca*^{H1047R}/Rosa26-YFP or control K14-rtTA/TetO-Cre/Rosa26-YFP induced for 4 months were sorted based on their Lin⁻YFP⁺CD29^{Lo}CD24⁺ or Lin⁻YFP⁺CD29^{Hi}CD24⁺ profiles. LCs were resuspended in 10 µl DMEM plus 50% bovine serum. BCs were sorted in the presence of 10 µM of Rock inhibitor (Y27632, Sigma) and resuspended in 75% DMEM/25% matrigel. Cell suspension was injected into the fourth mammary gland of 3-to-4-week-old NOD-SCID mice that had been cleared of endogenous epithelium as previously described^{18,19}. Recipient mice were mated 4 weeks after the transplantation, and killed 2-to-3 weeks later, when fully pregnant. Recipient glands were dissected and stained for GFP, K8 and K5 as whole mounts. An outgrowth was defined as an epithelial structure comprising ducts and lobules and/or terminal end buds.

Quantification of keratin⁺ cells within YFP⁺ cells. A total of 1,907, 1,704 and 2,391 YFP⁺ cells from three different mice per condition were analysed respectively in K8-CreER^{T2}/Rosa26-YFP induced 4 weeks, K8-CreER^{T2}/*Pik3ca*^{H1047R}/Rosa26-YFP induced 1 week and 8 weeks on 5 µm cryosections stained for K5, K8 and GFP. Coexpression of these markers was analysed with a confocal microscope. Cells were scored as K8⁺K5⁻ (K8), K8⁺K5⁺ (K5K8) or K8⁻K5⁺ (K5) and are shown in Fig. 3g.

Quantification of clone composition. Mammary glands were processed as whole mount and stained for K8, K5 and GFP. Clones were analysed by confocal microscopy. A total of 822, 936, 714 and 360 clones from three independent mice per condition were analysed in K8-CreER^{T2}/*Pik3ca*^{H1047R}/Rosa26-YFP induced for 1 week, induced for 10 weeks, and in K5-CreER^{T2}/*Pik3ca*^{H1047R}/Rosa26-YFP induced for 1 week and induced for 7 months respectively at dose of TAM that labelled very few and isolated cells. The clones were scored in three classes according to their keratin expression: luminal clones, composed only of K8⁺ cells, basal clones, composed only of K5⁺ cells, and mixed clones, composed of K8⁺ and K5⁺ cells. These data are shown in Extended Data Fig. 6p and Extended Data Fig. 7j.

Quantification of percentage YFP-labelled cells. The percentage of YFP-labelled cells within the luminal and basal populations was quantified by FACS. The luminal population was defined as the CD29^{Lo}CD24⁺ population and the basal population was defined as the CD29^{Hi}CD24⁺ population.

Whole-mount carmine staining. Whole-mount mammary fourth mammary glands were fixed in methanol Carnoy (60% methanol, 30% acetic acid, 10% chloroform) for at least 2 h and rehydrated in 70% ethanol, followed by water. Staining in carmine alum (Sigma) was done overnight and excess dye was rinsed with water. This is followed by incubation in 70%, 95%, 100% ethanol (1 h each) and fat-clearing in toluene overnight. All steps were carried out at room temperature.

Epithelial outgrowth measurement. Carmine-stained mammary glands were photographed with a Leica M80 stereomicroscope equipped with a Leica IC80 HD digital camera. The distance from the lymph node of the mammary epithelium was scored by measuring the distance between the distal edge of the lymph node and the most distal tip of the epithelium.

RNA extraction and quantitative real-time PCR. The protocol used for RNA extraction on FACS-isolated cells has been previously described³⁷. Briefly, RNA extraction was performed using the RNeasy micro kit (Qiagen) according to the manufacturer's recommendations and DNase treatment. After nanodrop RNA quantification and analysis of RNA integrity, purified RNA was used to synthesize the first-strand cDNA in a 50 µl final volume, using Superscript II (Invitrogen) and random hexamers (Roche). Genomic contamination was detected by performing the same procedure without reverse transcriptase. Quantitative PCR analyses were performed with 1 ng of cDNA as template, using FastStart Essential DNA green master (Roche) and a Light Cycler 96 (Roche) for real-time PCR system.

Relative quantitative RNA was normalized using the housekeeping gene *Gapdh*. Primers were designed using PrimerBank database (<http://pga.mgh.harvard.edu/primerbank/>) and are listed in Supplementary Table 4. Analysis of the results was performed using Light Cycler 96 software (Roche) and relative quantification was performed using the $\Delta\Delta C_t$ method using *Gapdh* as reference. The entire procedure was repeated in four biologically independent samples. For Extended Data Figs 6 and 7, data are shown as fold change over luminal cells or basal cells derived from 3-month-old wild-type mice (L-WT and B-WT).

Microarray analysis. Total RNA was analysed using mouse whole-genome MG-430 PM array from Affymetrix at the IRB Functional Genomics Core. All the results were normalized with RMA normalization using R-bioconductor package *affy* with standard parameters^{38,39}. Two biologically independent samples were analysed for each condition, except for tumours derived from K5-CreER^{T2}/*Pik3ca*^{H1047R}/Rosa26-YFP or K8-CreER^{T2}/*Pik3ca*^{H1047R}/Rosa26-YFP, for which three and seven samples were analysed, respectively. Sorted BCs from K5-CreER^{T2}/Rosa26-YFP mice induced for 8 weeks or 10–12 months, LCs from K8-CreER^{T2}/Rosa26-YFP mice induced for 8 weeks or 10–12 months, BCs and LCs from K5-CreER^{T2}/*Pik3ca*^{H1047R}/Rosa26-YFP mice induced for 10–12 months, BCs and LCs from K8-CreER^{T2}/*Pik3ca*^{H1047R}/Rosa26-YFP mice induced for 8 weeks, Lin⁻ cells from K5-CreER^{T2}/*Pik3ca*^{H1047R}/Rosa26-YFP-derived tumours and from K8-CreER^{T2}/*Pik3ca*^{H1047R}/Rosa26-YFP-derived tumours numbers 1, 2, 7, YFP⁺ cells from K8-CreER^{T2}/*Pik3ca*^{H1047R}/Rosa26-YFP-derived tumours numbers 3, 4, 5, 6 and from K8-CreER^{T2}/*Pik3ca*^{H1047R}/*p53*^{fl/fl}/Rosa26-YFP-derived tumours were analysed. The gene signature induced by the expression of PIK3CA(H1047R) in each population was determined by comparing their transcriptional profile with LCs arising from age-matched K8-CreER^{T2}/Rosa26-YFP or BCs arising from K5-CreER^{T2}/Rosa26-YFP mice. Only genes upregulated or downregulated by at least 1.5 fold were considered in the analysis.

Microarray data clustering. Clustering and bootstrap analyses were performed using the *pvc* and *gplots* packages of the R statistical suite⁴⁰. Clustering was performed with the default parameters of the *R* *hclust* function (Euclidean distance and complete linkage) considering only the top 500 most variant genes among all experiments.

Gene signature comparison. Venn diagrams were computed with the R statistical tool. The reported hypergeometric *P* values for every comparison between two signatures correspond to the probability to observe an intersection of at least a given size by chance only, knowing the number of genes tested on a microarray chip.

Murine and human breast tumours gene expression profile comparison. To compare the murine tumour gene expression data to human tumour data, we used the METABRIC data set composed of 1,992 patients⁴¹. METABRIC expression data were downloaded from the EBI website (data sets EGAD00010000210 and EGAD00010000211). When multiple probes mapped to the same Entrez gene identifier, we kept the one with the highest variance in the data set using the *genefu* package. The PAM50 subtypes were computed using the Bioconductor *genefu* package dedicated function³⁰ (1,448 basal, 1,027 HER2⁺, 2,260 LumB, 2,162 LumA and 323 normal).

Boxplots and Kruskal–Wallis test *P* values were computed using R. *P* values reflect the probability that at least one of the cancer subtypes express the tested signature at a significantly different level.

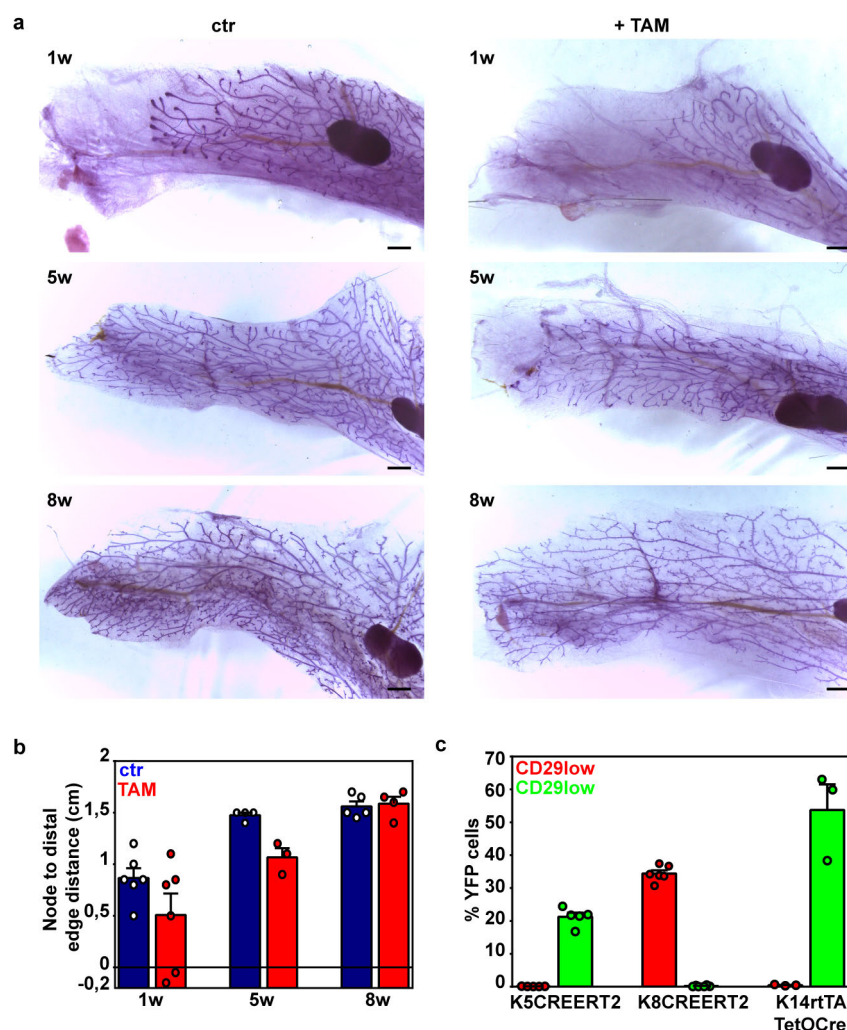
Uni-directional Student's *t*-test *P* values reflect the probability that one signature is significantly more expressed (or repressed) in one subtype compared to all the others. For the *t*-tests, as they are more robust to the extreme values, median and interquartile ranges were chosen as estimators of the central tendency and of the dispersion (instead of the mean and the standard deviation).

We then merged the murine data set with the METABRIC data set by keeping those genes described as orthologous in the Ensembl database downloaded via Biomart⁴² and having exactly the same identifier. Batch effect between murine and human data was corrected using the *Combat* function of the Bioconductor *sva* package⁴³. The PCA and clustering analyses were performed using the R statistical software considering an expression matrix containing only the expression values of the 46 PAM50 orthologous genes between mouse and human. For clustering, we used the Euclidean distance combined to the complete hierarchical clustering

method (default parameters). PAM50 subtypes were computed using the R/Bioconductor *genefu* package.

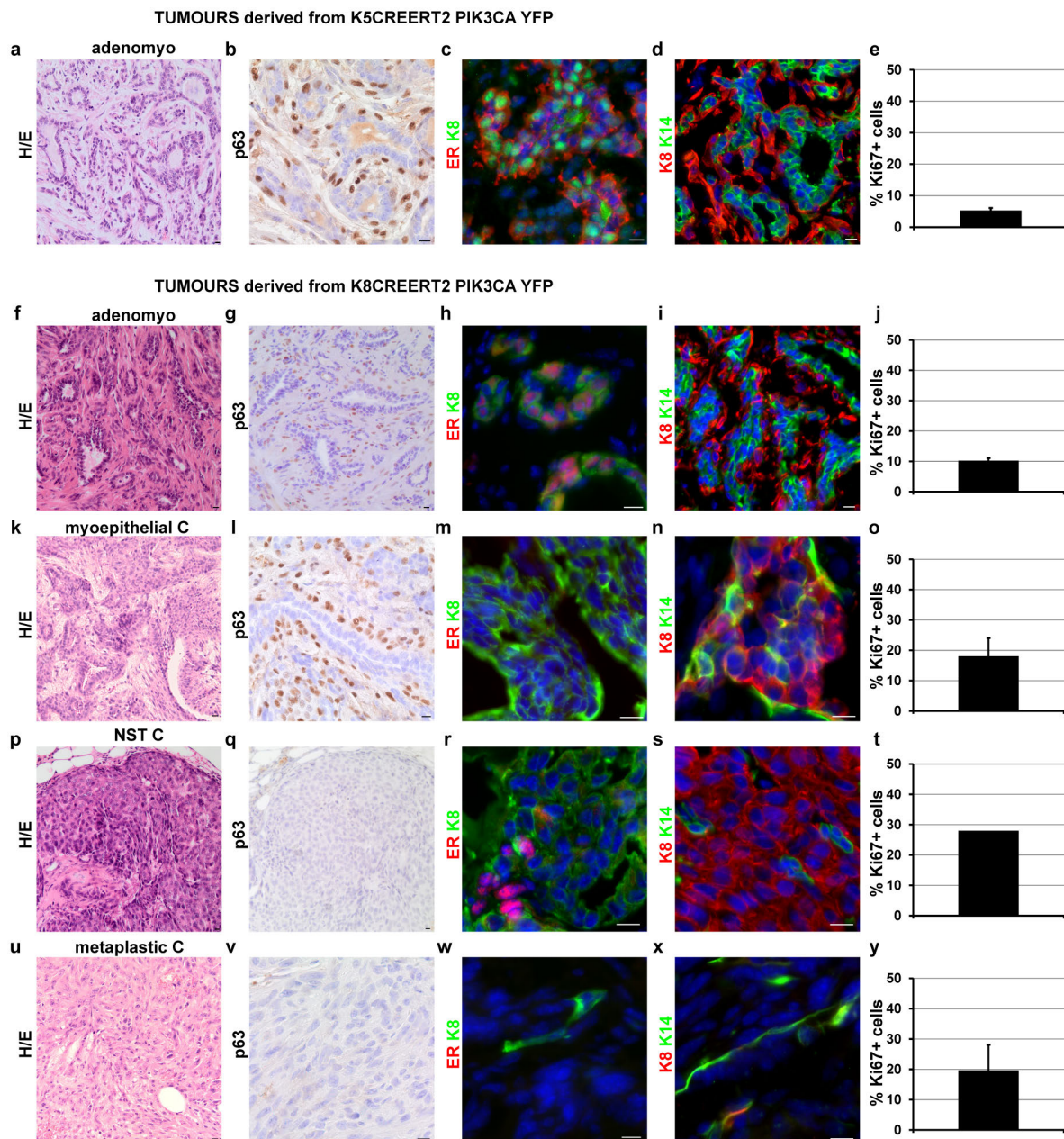
Survival analyses in humans. Mouse-derived signatures were converted to human signatures by considering the orthologous genes in humans. Signatures score were then computed and re-scaled using the dedicated function of the R/Bioconductor *genefu* package. The scores were computed for each patients of the METABRIC together with those present in 33 other breast tumours reference data sets³⁰ (7,220 patients). Survival curves were computed using the dedicated function of the *genefu* package only on untreated patients (1,859 cases) with available survival data. Expression level categories correspond to the tertiles of the expression values in the untreated patients. *P* values correspond to the log-rank *P* value, which reflects the probability that at least one of the class of signature expression presents a significantly differing outcome from the other classes.

31. Srinivas, S. *et al.* Cre reporter strains produced by targeted insertion of *EYFP* and *EGFP* into the *ROSA26* locus. *BMC Dev. Biol.* **1**, 4 (2001).
32. Nguyen, H., Rendl, M. & Fuchs, E. Tcf3 governs stem cell features and represses cell fate determination in skin. *Cell* **127**, 171–183 (2006).
33. Perl, A. K., Wert, S. E., Nagy, A., Lobe, C. G. & Whitsett, J. A. Early restriction of peripheral and proximal cell lineages during formation of the lung. *Proc. Natl Acad. Sci. USA* **99**, 10482–10487 (2002).
34. Jonkers, J. *et al.* Synergistic tumor suppressor activity of BRCA2 and p53 in a conditional mouse model for breast cancer. *Nature Genet.* **29**, 418–425 (2001).
35. Shehata, M., van Amerongen, R., Zeeman, A. L., Giraddi, R. R. & Stingl, J. The influence of tamoxifen on normal mouse mammary gland homeostasis. *Breast Cancer Res.* **16**, 411 (2014).
36. Rorive, S. *et al.* TIMP-4 and CD63: new prognostic biomarkers in human astrocytomas. *Mod. Pathol.* **23**, 1418–1428 (2010).
37. Beck, B. *et al.* A vascular niche and a VEGF-Nrp1 loop regulate the initiation and stemness of skin tumours. *Nature* **478**, 399–403 (2011).
38. McCall, M. N., Bolstad, B. M. & Irizarry, R. A. Frozen robust multiarray analysis (fRMA). *Biostatistics* **11**, 242–253 (2010).
39. Gentleman, R. C. *et al.* Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol.* **5**, R80 (2004).
40. Suzuki, R. & Shimodaira, H. Pvcust: an R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics* **22**, 1540–1542 (2006).
41. Curtis, C. *et al.* The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature* **486**, 346–352 (2012).
42. Kasperzyk, A. BioMart: driving a paradigm change in biological data management. *Database* **2011**, bar049 (2011).
43. Leek, J. T., Johnson, W. E., Parker, H. S., Jaffe, A. E. & Storey, J. D. The sva package for removing batch effects and other unwanted variation in high-throughput experiments. *Bioinformatics* **28**, 882–883 (2012).



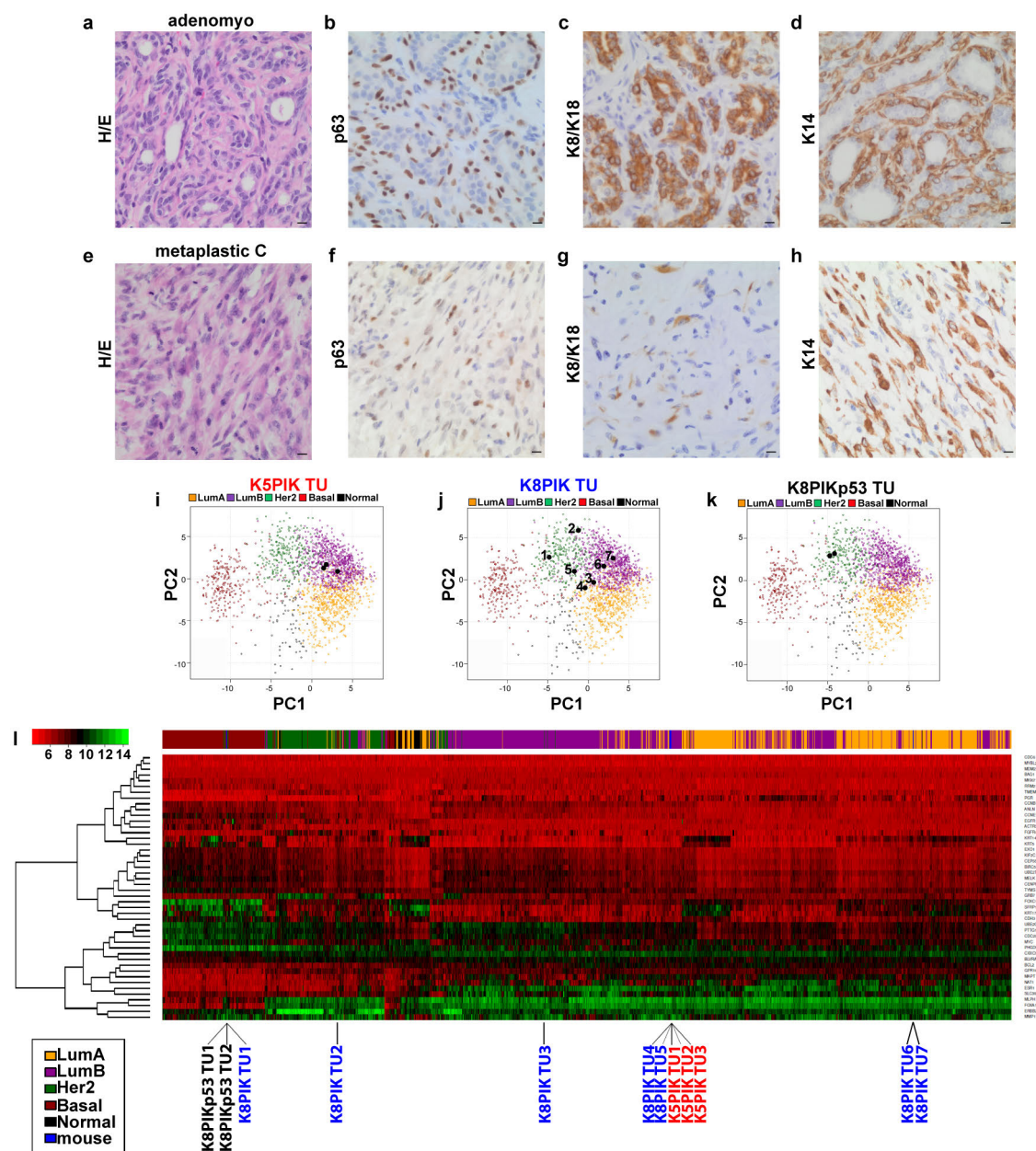
Extended Data Figure 1 | Tamoxifen administration has no long-term effect on the mammary gland. a, b, Effect of TAM on mammary epithelial postnatal growth. **a, b,** Representative whole-mount preparations of carmine alum-stained mammary epithelium from the fourth mammary gland, showing that TAM induces a delay in mammary epithelium growth at early time points, but no difference is observed 8 weeks after TAM induction (**a**) and mean distance from lymph node distal edge to the distal epithelial edge 1 week, 5 weeks and 8 weeks after TAM injection or oil injection (**b**) ($n = 6, 6, 4, 3, 5, 4$ mice respectively for 1 week control (ctr), 1 week TAM, 5 weeks control, 5 weeks TAM, 8 weeks control, 8 weeks TAM). P value derived from two-sided

Student's t -test is 0.161, 0.035, 0.748 when comparing control and TAM conditions at 1 week, 5 weeks and 8 weeks, respectively. **c,** Percentage of YFP⁺ cells in LCs (CD29^{low}/CD24⁺) and in BCs (CD29^{high}/CD24⁺) analysed by FACS 48 h after TAM administration in K5-CreER^{T2}/*Pik3ca*^{H1047R}/*Rosa26*-YFP and K8-CreER^{T2}/*Pik3ca*^{H1047R}/*Rosa26*-YFP, or 1 week after doxycycline administration to K14-rtTA/TetO-Cre/*Pik3ca*^{H1047R}/*p53*^{fl/+}/*Rosa26*-YFP mice ($n = 5, 6, 3$ mice respectively for K5-CreER^{T2}, K8-CreER^{T2} and K14-rtTA/TetO-Cre). Circles, individual data points. Scale bars, 100 μ m. Error bars, s.e.m.



Extended Data Figure 2 | Characterization of tumours derived from basal or luminal cells upon oncogenic *Pik3ca* expression. **a–e**, Characterization of adenomyoepithelioma (adenomyo) tumours derived from K5-CreER^{T2}/*Pik3ca*^{H1047R}/*Rosa26*-YFP mice. **f–y**, Characterization of tumours derived from K8-CreER^{T2}/*Pik3ca*^{H1047R}/*Rosa26*-YFP mice. **f–j**, Characterization of adenomyoepithelioma. **k–o**, Characterization of myoepithelial carcinoma (C). **p–t**, Characterization of invasive carcinoma of no special type (NST C).

u–y, Characterization of metaplastic carcinoma. **a, f, k, p, u**, Haematoxylin and eosin staining. **b, g, l, q, v**, p63 immunohistochemistry. **c, h, m, r, w**, Immunofluorescence of ER/K8. **d, i, n, s, x**, Immunofluorescence of K8/K14. **e, j, o, t, y**, Mean percentage of Ki67⁺ cells within tumours ($n = 6, 3, 3, 1, 3$ tumours and total number of cells counted = 10,408, 10,758, 11,174, 4,622, 5,732 in **e, j, o, t, y**, respectively). Error bars, s.e.m. Scale bars, 10 μ m.



Histological classification of tumours shown

K5PIK TU1 adenomyoepithelioma

K5PIK TU2 adenomyoepithelioma

K5PIK TU3 adenomyoepithelioma

K8PIK TU1 myoepithelial carcinoma

K8PIK TU2 invasive carcinoma NST

K8PIK TU3 metaplastic carcinoma with mesenchymal differentiation

K8PIK TU4 adenomyoepithelioma

K8PIK TU5 adenomyoepithelioma

K8PIK TU6 adenomyoepithelioma+ myoepithelial carcinoma

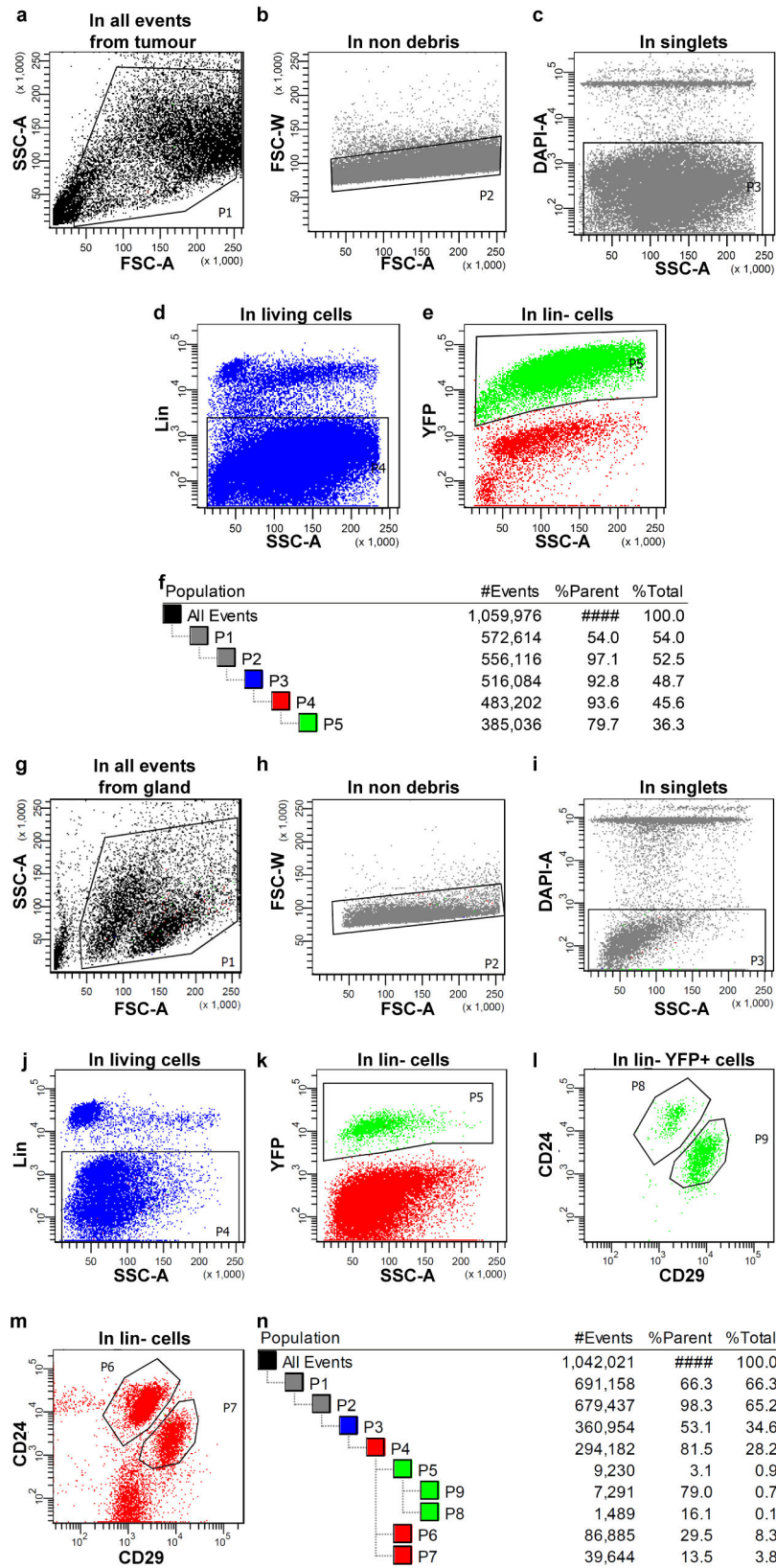
K8PIK TU7 adenomyoepithelioma

K8PIKp53 TU1 metaplastic carcinoma with mesenchymal differentiation

K8PIKp53 TU2 metaplastic carcinoma with mesenchymal differentiation

Extended Data Figure 3 | Similarities between mouse *Pik3ca*-derived mammary tumours and human breast cancers. **a–d**, Human breast tumour histologically classified as adenomyoepithelioma resembling K5-CreER^{T2}/*Pik3ca*^{H1047R}/Rosa26-YFP-derived tumours (K5PIK TU). **a**, Haematoxylin and eosin staining. **b–d**, p63 (**b**), K8/K18 (**c**) and K14 (**d**) immunohistochemistry in the human adenomyoepithelioma. **e–h**, Human breast tumour histologically classified as metaplastic carcinoma resembling K8-CreER^{T2}/*Pik3ca*^{H1047R}/Rosa26-YFP derived tumours (K8PIK TU). **e**, Haematoxylin and eosin staining. **f–h**, p63 (**f**), K8/K18 (**g**) and K14 (**h**) immunohistochemistry in the human metaplastic carcinoma. **i–k**, Principal component analysis (PCA) of the METABRIC patients together with murine tumours according to the expression values of the PAM50 genes common to mice and humans. **i**, PCA of three K5-CreER^{T2}/*Pik3ca*^{H1047R} tumours (black dots) showing that these

tumours cluster with human luminal B cancer subtype. **j**, PCA of seven K8-CreER^{T2}/*Pik3ca*^{H1047R}/Rosa26-YFP-derived tumours (numbered black dots). Histological classification of each numbered tumour is described below the figure. **k**, PCA of two K8-CreER^{T2}/*Pik3ca*^{H1047R}/*p53*^{fl/fl}/Rosa26-YFP-derived tumours (K8PIKp53 TU) (black dots) showing that these tumours cluster together with human HER2⁺ subtype. **l**, Clustering of the murine tumours among human tumours of the METABRIC data set. Clustering was performed by grouping tumours presenting similar expression patterns of PAM50 genes. Colours on top of the heatmap represent the PAM50 subtypes attributed to the human tumours. The discrepancy between PCA and clustering analysis are due to the influence of HER2 low expression in these tumours, for which around 60% of PC2 relies on ERBB2 expression. Scale bars, 10 μ m.

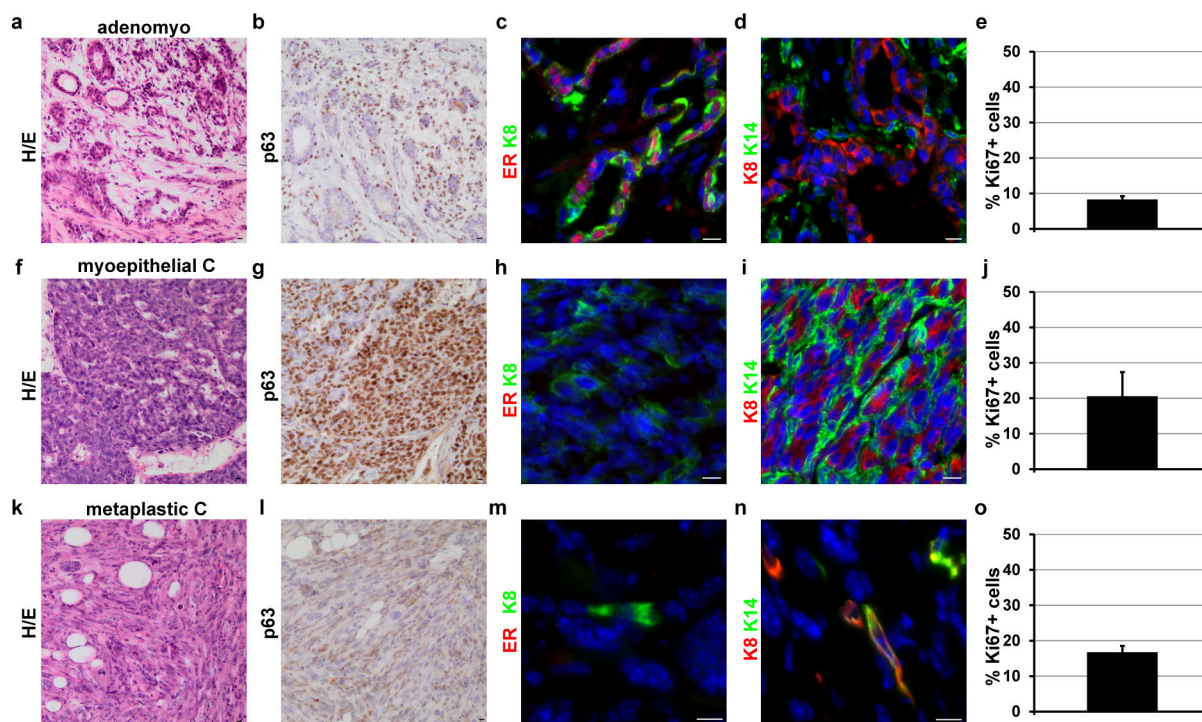


Extended Data Figure 4 | Gating strategy to analyse and isolate tumour cells, LCs, and BCs according to their YFP, CD29 and CD24 profile.

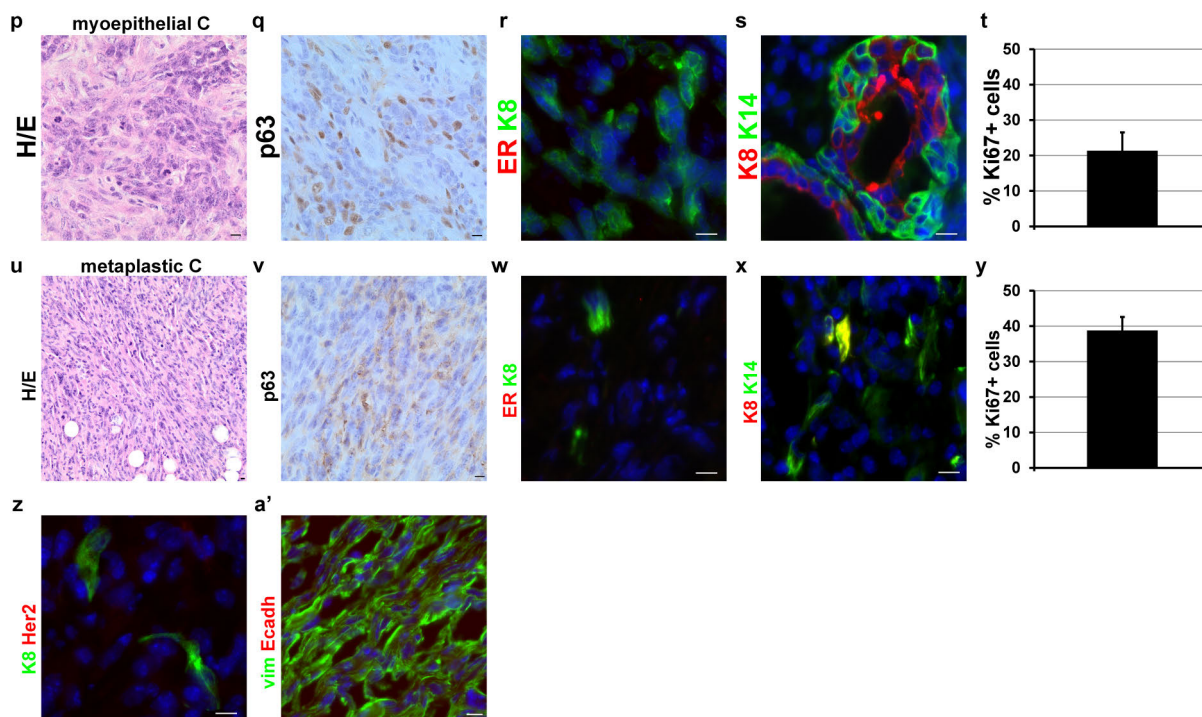
a–e, Dot plot FACS analysis of unicellular suspension of mammary tumour cells (in this example from K8-CreER^{T2}/*Pik3ca*-H1047R/Rosa26-YFP tumour) stained for Lin (CD31, CD45, CD140a). Debris were eliminated from all events in P1 (**a**), doublets were discarded in P2 (**b**), the living cells were gated in P3 by DAPI dye exclusion (**c**), the non-epithelial Lin⁺ cells were discarded in P4 (**d**), and the YFP⁺ cells were gated in P5 (**e**). **f**, Gating strategy used for FACS analysis and cell sorting, showing the proportion of parent and total cells for each gate. Tumour cells were isolated based on their Lin[−] profile for YFP[−] tumours (P4 gate), or were isolated based on their YFP profile (P5 gate) for the YFP⁺ tumours, as described in Methods. **g–m**, Dot plot FACS

analysis of unicellular suspension of mammary cells (in this example from K5-CreER^{T2}/*Pik3ca*^{H1047R}/Rosa26-YFP mice 12 months after TAM induction) stained for CD24, CD29 and Lin (CD31, CD45, CD140a). Debris were eliminated from all events in P1 (**g**), doublets were discarded in P2 (**h**), the living cells were gated in P3 by DAPI dye exclusion (**i**), the non-epithelial Lin⁺ cells were discarded in P4 (**j**), and the YFP⁺ cells were gated in P5 (**k**). **l, m**, CD29 and CD24 expression were used to gate the CD29^{Lo}CD24⁺ population, corresponding to LCs, and to gate the CD29^{Hi}CD24⁺ population, corresponding to BCs, either in YFP⁺ cells (**l**) or in Lin[−] cells (**m**). **n**, Gating tree showing the gating strategy used for FACS analysis and sorting, showing the proportion of parent and total cells for each gate.

TUMOURS derived from K14rtTA TetOCre PIK3CA p53fl/+ YFP

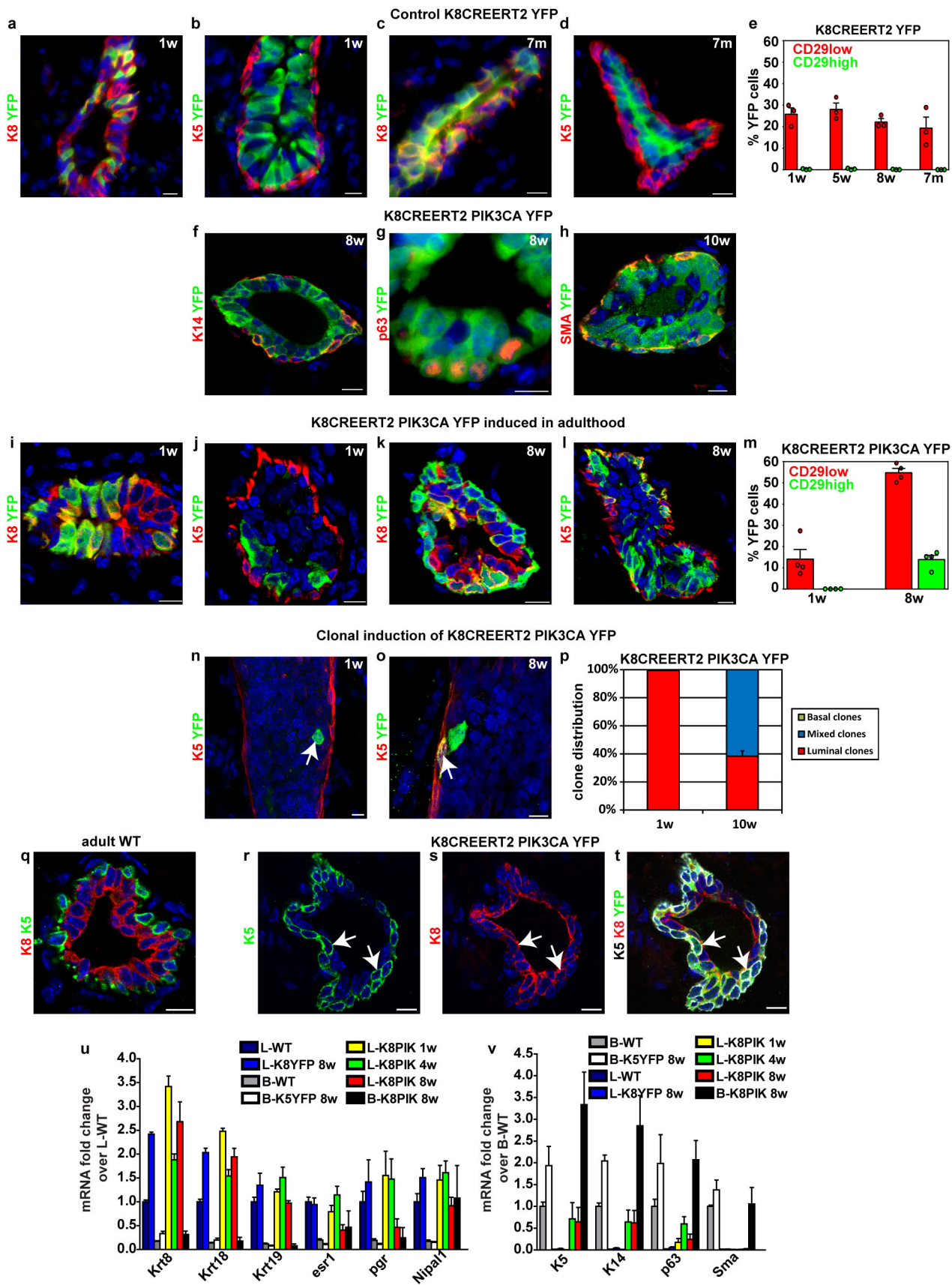


TUMOURS derived from K8CREERT2 PIK3CA p53fl/fl YFP



Extended Data Figure 5 | Characterization of tumours derived from BCs or LCs upon concomitant expression of oncogenic *Pik3ca* and deletion of *p53*. **a–o,** Characterization of tumours derived from K14-rtTA/TetO-Cre/*Pik3ca*^{H1047R}/*p53*^{fl/+}/Rosa26-YFP mice. **a–e,** Characterization of adenomyo-epithelioma (adenomyo). **f–j,** Characterization of myoepithelial carcinoma. **k–o,** Characterization of metaplastic carcinoma. **p–a',** Characterization of tumours derived from K8-CreER^{T2}/*Pik3ca*^{H1047R}/*p53*^{fl/fl}/Rosa26-YFP mice. **p–t,** Characterization of myoepithelial carcinoma (C). **u–a',** Characterization

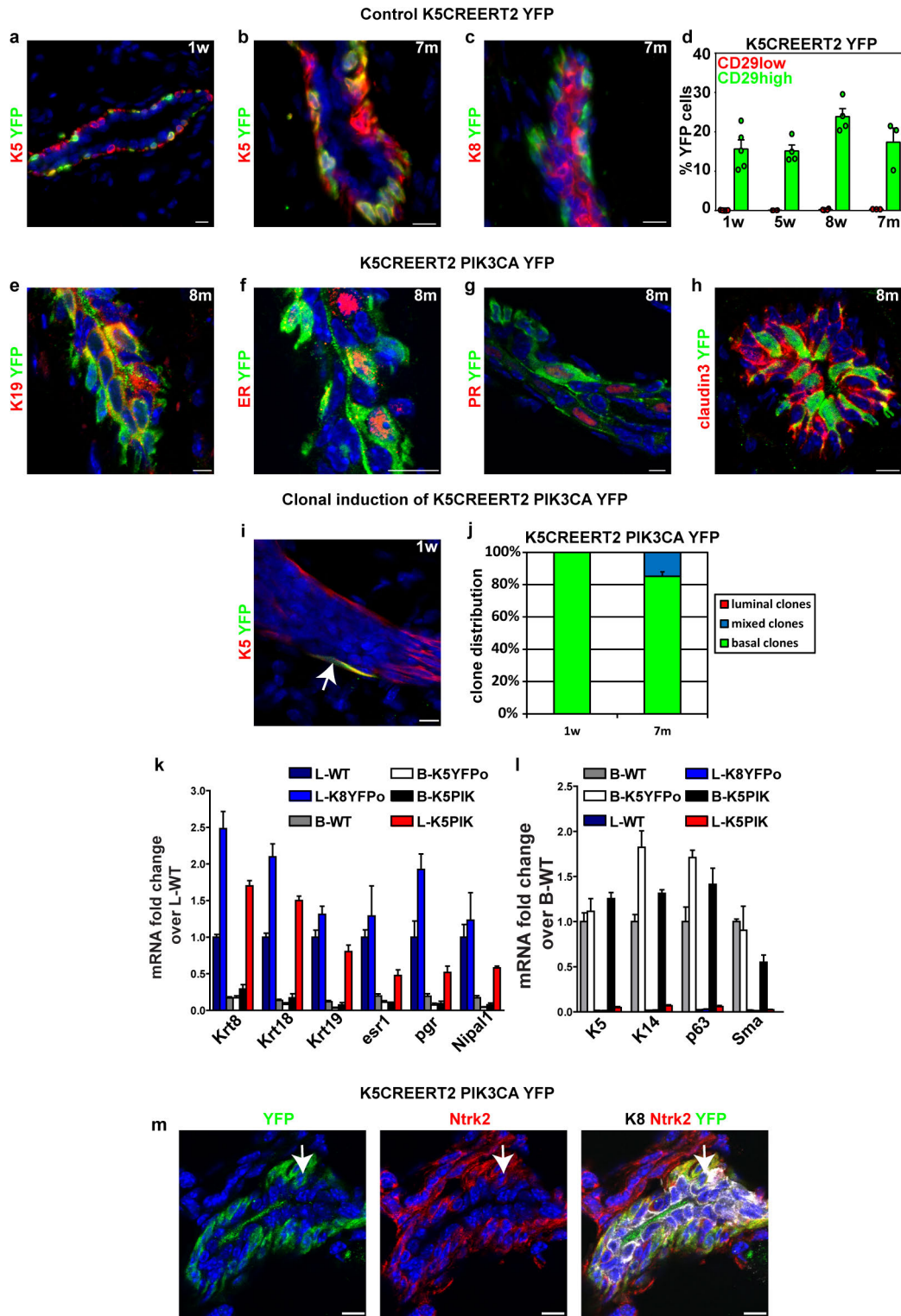
of metaplastic carcinoma. **a, f, k, p, u,** Haematoxylin and eosin staining. **b, g, l, q, v,** p63 immunohistochemistry. **c, h, m, r, w,** Immunofluorescence of K8/ER. **d, i, n, s, x,** Immunofluorescence of K8/Ki14. **e, j, o, t, y,** Mean percentage of Ki67⁺ cells within tumours ($n = 4, 3, 3, 4, 6$ tumours and total cells counted = 11,903, 10,670, 6,992, 14,743, 8,172 in **e, j, o, t, y**, respectively). **z,** Immunofluorescence of K8/HER2. **a',** Immunofluorescence of E-cadherin/vimentin. Error bars, s.e.m. Scale bars, 10 μ m.



Extended Data Figure 6 | Oncogenic *Pik3ca* expression induces

multipotency in unipotent luminal progenitors. **a–d**, Immunofluorescence showing the expression of K8/YFP (**a**, **c**) or K5/YFP (**b**, **d**) 1 week (**a**, **b**) and 7 months (**c**, **d**) after TAM injection in control K8-CreER^{T2}/Rosa26-YFP mammary gland. **e**, Percentage of YFP⁺ cells in LCs (CD29^{Lo}/CD24⁺) and in BCs (CD29^{Hi}/CD24⁺) at different time points after TAM administration to K8-CreER^{T2}/Rosa26-YFP mice ($n = 3$ mice per time point) showing that no YFP⁺ cells expressing CD29^{Hi}/CD24⁺ were detected in control K8-CreER^{T2}/Rosa26-YFP mammary glands at any time point. **f–h**, Immunofluorescence of K14/YFP (**f**), p63/YFP (**g**), SMA/YFP (**h**) 8 weeks (**f**, **g**) or 10 weeks (**h**) after TAM administration to K8-CreER^{T2}/*Pik3ca*^{H1047R}/Rosa26-YFP mice, shows that the BCs arising from LCs upon oncogenic *Pik3ca* targeting expressed these classical markers of BCs. **i–m**, Induction of *Pik3ca*^{H1047R} expression in LCs in adult mice. **i–l**, Immunofluorescence showing the expression of K8/YFP (**i**, **k**) or K5/YFP (**j**, **l**) 1 week (**i**, **j**) and 8 weeks (**k**, **l**) after TAM injection in K8-CreER^{T2}/*Pik3ca*^{H1047R}/Rosa26-YFP mice induced in adulthood. **m**, Percentage of YFP⁺ cells in LCs (CD29^{Lo}/CD24⁺) and in BCs (CD29^{Hi}/CD24⁺) at different time points after TAM administration to K8-CreER^{T2}/*Pik3ca*^{H1047R}/Rosa26-YFP mice induced in adulthood ($n = 4$ mice per time point). **n**, **o**, Immunofluorescence of K5/YFP showing the clonal YFP expression in a single isolated LC 1 week after TAM injection (**n**), and

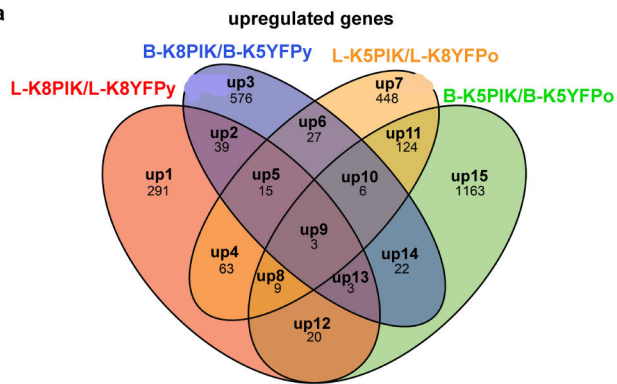
8 weeks after TAM injection showing a clone that gave rise to an LC and a BC (**o**) in K8-CreER^{T2}/*Pik3ca*^{H1047R}/Rosa26-YFP mammary gland. Arrow in **n** points to the isolated LC, while arrow in **o** points to the newly arisen BC. **p**, Distribution of clones 1 week or 10 weeks after TAM injection in K8-CreER^{T2}/*Pik3ca*^{H1047R}/Rosa26-YFP at clonal dose. Clones were scored as composed of only luminal cells (luminal clones), composed of only basal cells (basal clones) or composed of luminal and basal cells (mixed clones) ($n = 4$ mice per time point). See Methods for more details. **q–t**, Immunofluorescence of K5/K8 (**q**), K5 (**r**), K8 (**s**) and K5/K8/YFP (**t**) shows that in wild-type mammary gland, K5 and K8 are not co-expressed (**q**), while K5/K8 double-positive cells are observed in K8-CreER^{T2}/*Pik3ca*^{H1047R}/Rosa26-YFP mammary gland 8 weeks after oncogenic *Pik3ca* expression in LCs (**r–t**). Arrows in **r–t** point to K5⁺K8⁺YFP⁺ cells. **u**, **v**, RT-PCR analysis of luminal (**u**) or basal (**v**) genes in YFP⁺ LCs and BCs sorted from K8-CreER^{T2}/*Pik3ca*^{H1047R}/Rosa26-YFP mice induced for 1 week, 4 weeks or 8 weeks, in YFP⁺ LCs derived from K8-CreER^{T2}/Rosa26-YFP and in YFP⁺ BCs derived from K5-CreER^{T2}/Rosa26-YFP mice induced for 8 weeks. Data for luminal genes are compared to adult wild-type LCs (**u**) while data for basal genes are compared to adult wild-type BCs (**v**) ($n = 4$ biologically independent samples per condition). Circles, individual data points. Scale bars, 10 μ m. Error bars, s.e.m.



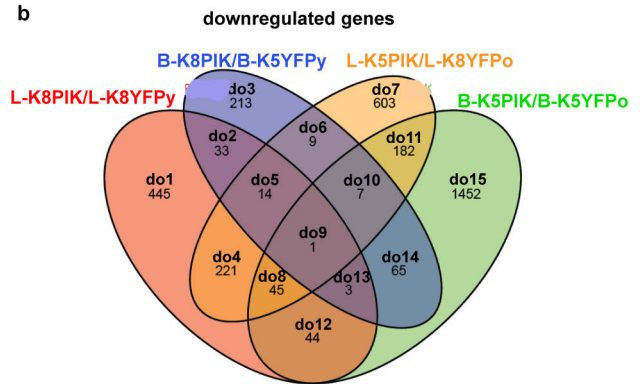
Extended Data Figure 7 | Oncogenic *Pik3ca* expression induces multipotency in unipotent basal progenitors. **a–c**, Immunofluorescence showing the expression of K5/YFP (**a**, **b**) or K8/YFP (**c**) at 1 week (**a**) and 7 months (**b**, **c**) in control K5-CreER^{T2}/Rosa26-YFP mammary gland. **d**, Percentage of YFP⁺ cells in LCs (CD29^{Lo}/CD24⁺) and in BCs (CD29^{Hi}/CD24⁺) at different time points after TAM administration to K5-CreER^{T2}/Rosa26-YFP ($n = 5, 4, 4, 3$ mice for 1 week, 8 weeks, 7 months and 12 months, respectively) showing that no YFP⁺ cells expressing CD29^{Lo}/CD24⁺ were detected in control K5-CreER^{T2}/Rosa26-YFP mammary glands at any time point. **e–h**, Immunofluorescence of K19/YFP (**e**), ER/YFP (**f**), PR/YFP (**g**), claudin 3/YFP (**h**), 8 months after TAM administration to K5-CreER^{T2}/*Pik3ca*^{H1047R}/Rosa26-YFP mice, shows that LCs arising from BCs upon oncogenic *Pik3ca* targeting expressed these classical markers of LCs. **i**, Immunofluorescence of K5/YFP showing the YFP expression in a single isolated BC 1 week after TAM injection at a clonal dose. Arrow points to the isolated BC. **j**, Distribution of clones 1 week or 7 months after TAM injection in K5-CreER^{T2}/*Pik3ca*^{H1047R}/Rosa26-YFP at

a clonal dose. Clones were scored as composed of only luminal cells (luminal clones), composed of only basal cells (basal clones) or composed of luminal and basal cells (mixed clones) ($n = 3, 4$ mice for 1 week and 7 months, respectively). See Methods for more details. **k**, **l**, RT-PCR analysis of luminal (**k**) or basal (**l**) genes in YFP⁺ LCs and BCs sorted from K5-CreER^{T2}/*Pik3ca*^{H1047R}/Rosa26-YFP mice induced for 10–12 months, in YFP⁺ LCs derived from K8-CreER^{T2}/Rosa26-YFP mice and in YFP⁺ BCs derived from K5-CreER^{T2}/Rosa26-YFP mice induced for 10–12 months. Data for luminal genes are compared to adult wild-type LCs (**k**) while data for basal genes are compared to adult wild-type BCs (**l**) ($n = 4$ biologically independent samples per condition). **m**, Confocal microscopy analysis of immunofluorescence of YFP, Ntrk2 and K8 of mammary glands 7 months after *Pik3ca* expression in BCs, showing that the newly formed LCs after *Pik3ca* expression in BCs co-expressed Ntrk2 and K8. Arrow points to formed K8⁺/Ntrk2⁺/YFP⁺ cell. Circles, individual data points. Scale bars, 10 μ m. Error bars, s.e.m.

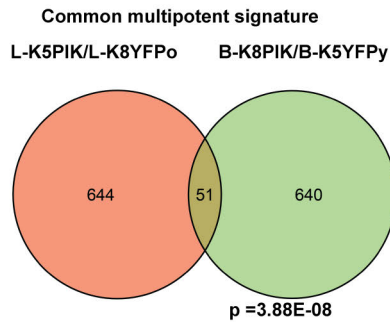
a



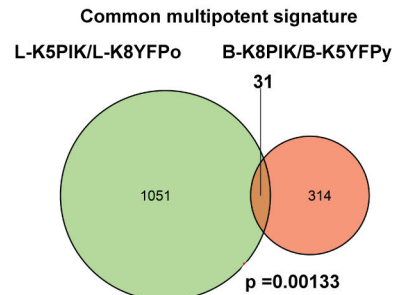
b



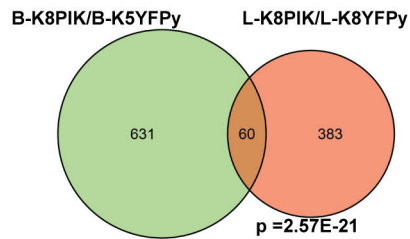
c



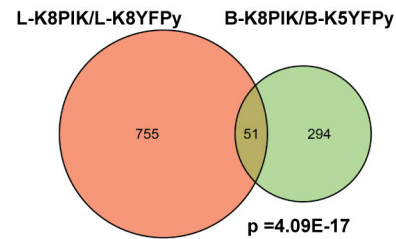
d



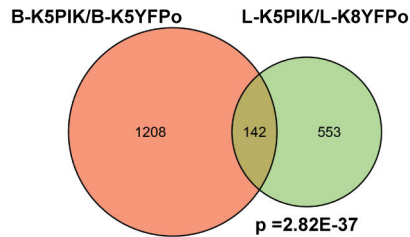
e



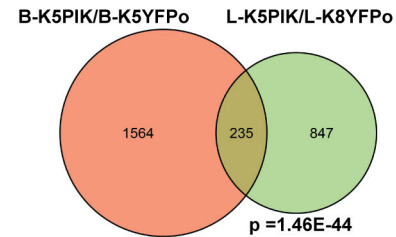
f



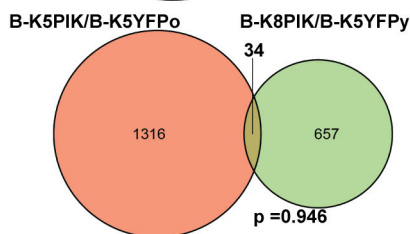
g



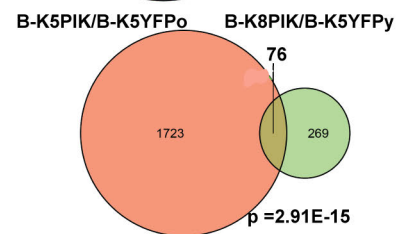
h



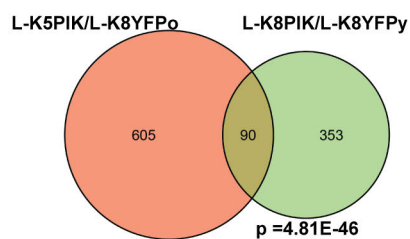
i



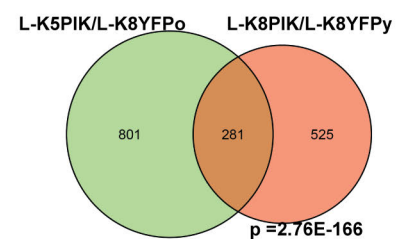
j



k

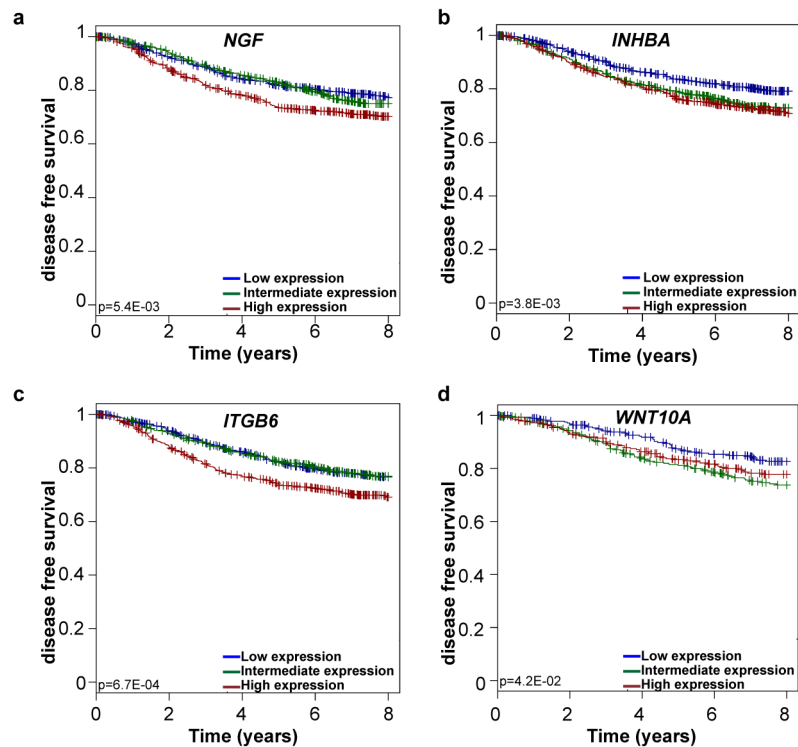


l



Extended Data Figure 8 | Molecular characterization of oncogenic *Pik3ca* induced multipotency. **a, b,** Venn diagram representing the common and distinct upregulated (**a**) and downregulated (**b**) genes in BCs and LCs after *Pik3ca* expression in BCs and LCs compared to age-matched control BCs and LCs, respectively, with the name of the list of genes and number of genes in each section. The list of genes in each Venn section is provided in Supplementary Tables 2 and 3. **c–l,** Venn diagrams representing the common genes upregulated (**c, e, g, i, k**) or downregulated (**d, f, h, j, l**) in the newly generated LCs or BCs after *Pik3ca*^{H1047R} expression in unipotent progenitors (**c, d**); in LCs and in BCs after *Pik3ca*^{H1047R} expression in LCs (genes regulated following the initial targeting of *Pik3ca*^{H1047R} in LCs, and thus reflecting

the LC of origin) (**e, f**); in LCs and in BCs after *Pik3ca*^{H1047R} expression in BCs (genes regulated by *Pik3ca*^{H1047R} in BCs, and thus reflecting the BC of origin) (**g, h**); in BCs after *Pik3ca*^{H1047R} expression in LCs and in BCs (genes regulated by *Pik3ca*^{H1047R} expression in BCs, irrespective of cell of origin) (**i, j**); in LCs after *Pik3ca*^{H1047R} expression in BCs and in LCs (genes regulated by *Pik3ca*^{H1047R} expression in LCs, irrespective of cell of origin) (**k, l**). Diameter of the diagram is proportional to the number of genes it contains. The reported hypergeometric *P* values correspond to the probability of observing an intersection of this size by chance only, knowing the number of genes tested on a microarray chip.



Extended Data Figure 9 | Genes of luminal-to-basal multipotency signature correlate with patient outcome in untreated breast cancer patients. a–d, Disease-free survival in untreated patients according to the level of expression (low = blue, intermediate = green or high = red) of the genes in the luminal-to-basal multipotency signature, namely *NGF* (a), *INHBA* (b), *ITGB6* (c) and *WNT10A* (d), showing that genes of luminal-to-basal

multipotency signature predict disease-free survival in untreated breast cancer patients. Patients expressing high levels of this signature are more prone to tumour relapse while those expressing lower levels of this signature show lower rates of relapse. The log-rank *P* values account for the significance of this difference.

Regulation of mitochondrial morphology and function by stearylolation of TFR1

Deniz Senyilmaz¹, Sam Virtue², Xiaojun Xu^{1†}, Chong Yew Tan², Julian L. Griffin³, Aubry K. Miller¹, Antonio Vidal-Puig^{2,4} & Aurelio A. Teleman¹

Mitochondria are involved in a variety of cellular functions, including ATP production, amino acid and lipid biogenesis and breakdown, signalling and apoptosis^{1–3}. Mitochondrial dysfunction has been linked to neurodegenerative diseases, cancer and ageing⁴. Although transcriptional mechanisms that regulate mitochondrial abundance are known⁵, comparatively little is known about how mitochondrial function is regulated. Here we identify the metabolite stearic acid (C18:0) and human transferrin receptor 1 (TFR1; also known as TFR1) as mitochondrial regulators. We elucidate a signalling pathway whereby C18:0 stearylates TFR1, thereby inhibiting its activation of JNK signalling. This leads to reduced ubiquitination of mitofusin via HUWE1, thereby promoting mitochondrial fusion and function. We find that animal cells are poised to respond to both increases and decreases in C18:0 levels, with increased C18:0 dietary intake boosting mitochondrial fusion *in vivo*. Intriguingly, dietary C18:0 supplementation can counteract the mitochondrial dysfunction caused by genetic defects such as loss of the Parkinson's disease genes *Pink* or *Parkin* in *Drosophila*. This work identifies the metabolite C18:0 as a signalling molecule regulating mitochondrial function in response to diet.

To study the function of very long chain fatty acids, we analysed *Drosophila* lacking *Elovl6* (refs 6, 7), the enzyme elongating C16 fatty acids to C18. Sequence analysis identified *noa*⁸ as fly *Elovl6* (herein referred to as *Elovl6*). On standard laboratory food, *Elovl6* loss-of-function animals (*l(3)02281*^{–/–}; *Elovl6*[–]) die as early larvae⁸ (Fig. 1a). We confirmed that *Elovl6*[–] mutants have impaired C16:0→C18:0 elongase activity and reduced C18:0 levels (Extended Data Fig. 1a, b), and that their lethality is rescued by human *ELOVL6* (Extended Data Fig. 1c, d). Survival to pupation was rescued by supplementing fly food (containing little lipid), with C18:0 (Fig. 1a), but not C18:1 or C20:0 (Extended Data Fig. 1e), confirming that the larval lethality is due to C18:0 deficit.

We serendipitously discovered that removing antifungal agents from fly food improved survival of *Elovl6*[–] mutants (Fig. 1a). Since these agents are mitotoxins, this suggested that *Elovl6*[–] mutants might be hypersensitive to mitochondrial inhibition. Indeed, sub-lethal concentrations of rotenone, a mitochondrial respiratory chain complex I inhibitor, killed *Elovl6*[–] mutants when added to antifungal-free food (Fig. 1b), but other drugs did not (Extended Data Fig. 1f). Thus, mitochondrial function is limiting in *Elovl6*[–] mutants. *Elovl6*[–] mutants have impaired mitochondrial respiration, rescued by dietary C18:0 supplementation (Fig. 1c) or by expressing an alternative oxidase, AOX⁹, allowing bypass of complexes III and IV (Fig. 1d). Complex IV activity was not impaired in *Elovl6*[–] mutants (Extended Data Fig. 1g), suggesting that *Elovl6*[–] mutants suffer from a complex III defect.

If the main cause of *Elovl6*[–] lethality is reduced mitochondrial function, then viability should be rescued by restoring mitochondrial functional capacity. Indeed, *Elovl6*[–] viability was rescued by expressing AOX or *Spargel* (*Drosophila* *PGC1A*), driving mitochondrial

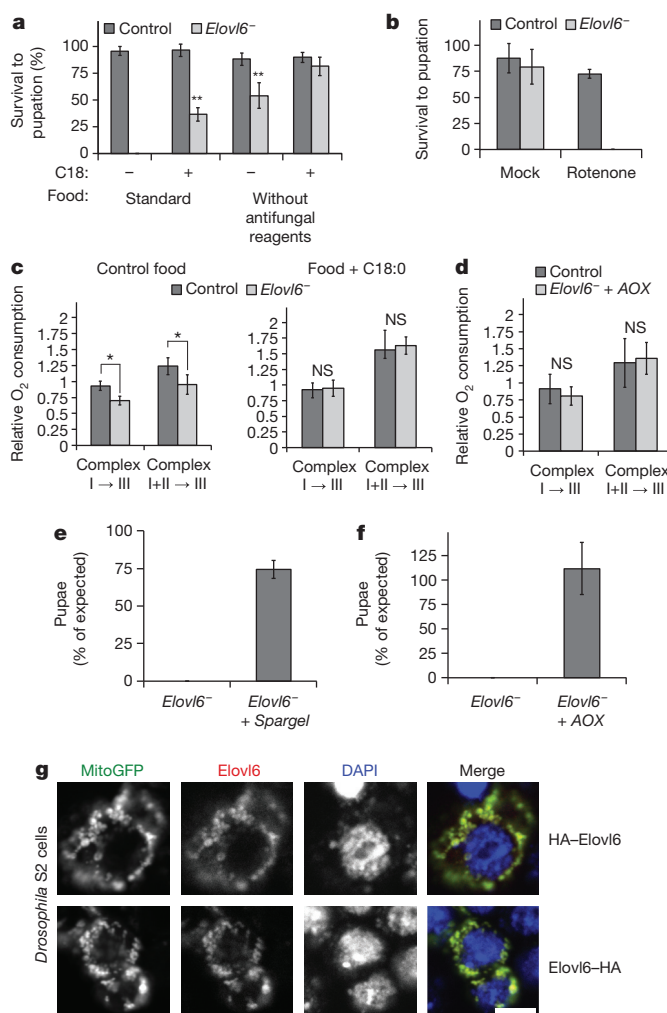


Figure 1 | *Drosophila* lacking C18:0 have impaired mitochondrial function. **a**, *Elovl6* mutant larval lethality rescued by dietary C18:0 (10% in food) or by removal of mitotoxic antifungal reagents ($n = 4 \times 60$ animals per vial). **b**, *Elovl6* mutants are sensitive to sub-lethal concentrations (100 μ M) of rotenone ($n = 4 \times 30$ animals per vial). **c**, **d**, *Elovl6* mutants have impaired respiration (c, left), rescued by supplementing food with C18:0 (10%) (c, right), or by expressing *Ciona intestinalis* alternative oxidase (AOX) (d), allowing bypass of complexes III and IV. $n = 4 \times 6$ animals. **e**, **f**, Survival to pupation of *Elovl6* mutants is rescued by ubiquitous expression of *Spargel* (e) or AOX (f). χ^2 tests, $P = 0.05$, $n = 195$ (e) or 81 (f). **g**, Amino- or carboxy-terminus-tagged *Drosophila* *Elovl6* localizes to mitochondria, visualized with mitoGFP in S2 cells. DAPI, 4',6-diamidino-2-phenylindole; HA, haemagglutinin. Scale bar, 10 μ m ($n = 4$). For details, see Supplementary Methods. Error bars show standard deviation (s.d.). **a–d**, ** $P < 0.01$, * $P < 0.05$, not significant (NS) $P > 0.05$, two-tailed t -test.

¹German Cancer Research Center (DKFZ), 69120 Heidelberg, Germany. ²University of Cambridge Metabolic Research Laboratories, Wellcome Trust-MRC Institute of Metabolic Science, Cambridge CB2 0QQ, UK. ³The Department of Biochemistry, Tennis Court Road, Cambridge CB2 1GA, UK. ⁴Wellcome Trust Sanger Institute, Hinxton, Cambridgeshire CB10 1SA, UK. [†]Present address: State Key laboratory of Natural Medicines, China Pharmaceutical University, 210009 Nanjing, China.

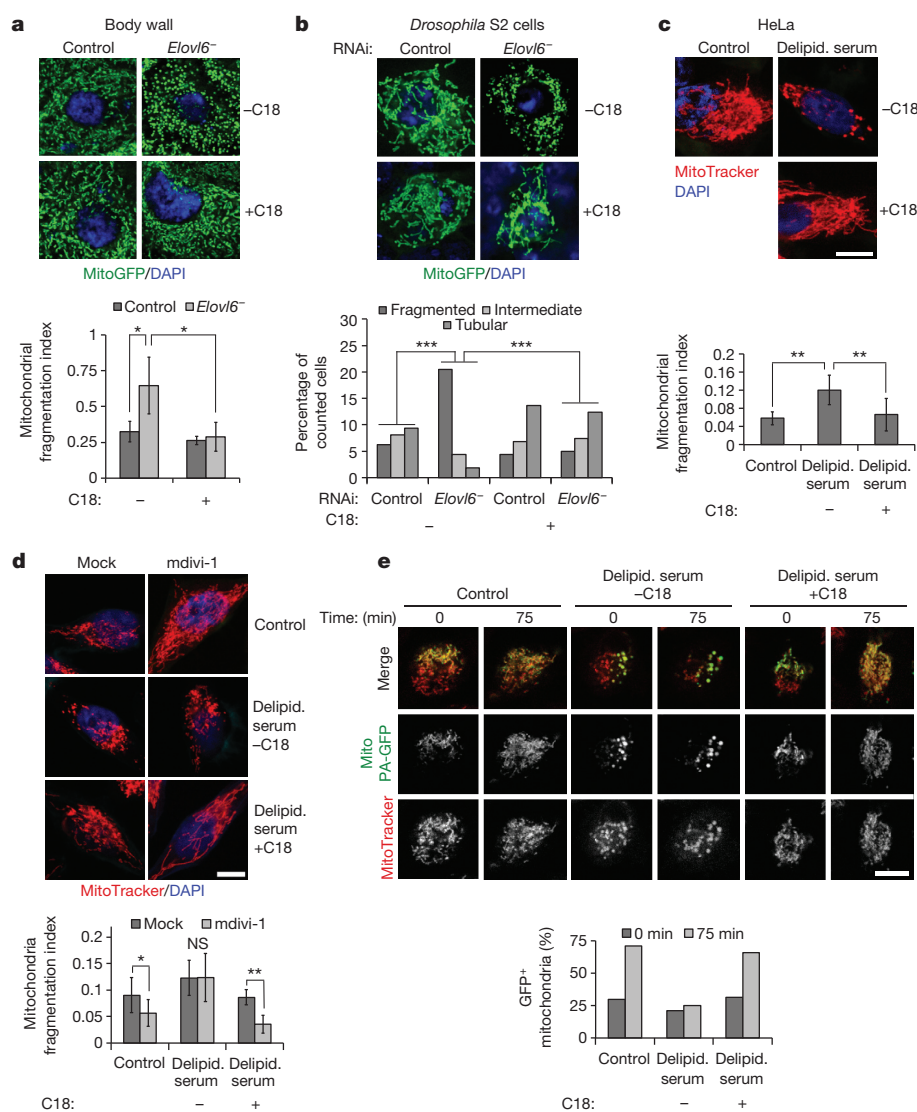


Figure 2 | C18:0 is required for mitochondrial fusion. **a**, *Elov6* mutants have fragmented mitochondria (top), rescued by dietary C18:0 (10% in food). Bottom, fragmentation quantified (8 fields from 4 animals). **b**, *Elov6* knockdown in *Drosophila* cells causes mitochondrial fragmentation, reversed by supplementing medium with 100 μ M C18:0 for 120 min. Bottom, quantification. RNAi, RNA interference. $n = 50$. *** $P < 0.001$, Mann–Whitney test. **c**, C18:0 removal by delipidating serum (Delipid. serum) causes mitochondrial fragmentation in human cells, rescued by resupplementing with 100 μ M C18:0 for 2 h. Bottom, quantification. $n = 15$. **d**, C18:0 affects mitochondrial fusion, not fission. Pharmacological inhibition of mitochondrial fission with mdivi-1 does not cause mitochondrial network fusion in cells growing without C18:0. Bottom, quantification. $n = 15$. **e**, Direct observation of mitochondrial fusion, monitored as dispersion of locally photoactivated mitoGFP (green) into the rest of the mitochondrial network (MitoTracker, red). Bottom, quantification (representative of ten experiments). For details, see Supplementary Methods. Error bars show s.d. **a**, **c**, **d**, bottom, * $P < 0.05$, ** $P < 0.01$, not significant (NS), $P \geq 0.05$ two tailed t -test. Scale bars: 20 μ m (**a**); 10 μ m (**b–e**).

biogenesis (Fig. 1e, f and Extended Data Fig. 1h, i). Thus, the organismal function of C18:0 is less pleiotropic than expected. Interestingly, *Drosophila* *Elov6* localizes to mitochondrial outer membranes (Fig. 1g and Extended Data Fig. 1j).

Lipidomic analysis of purified larval mitochondria (Extended Data Fig. 5b) revealed that their membranes have little C18:0 (Extended Data Fig. 2a), suggesting that C18:0 does not have a structural role in mitochondria. *Elov6*^{-/-} mutants also did not have fewer mitochondria than controls (assayed by porin levels and citrate synthase activity; Extended Data Fig. 2b, c). We therefore investigated whether C18:0 regulates mitochondrial activity. Mitochondria dynamically fuse and fission to form tubular structures^{10,11}. *Elov6*^{-/-} mutants had hyperfragmented mitochondria, rescued by dietary C18:0 supplementation (Fig. 2a and Extended Data Fig. 2d). *Elov6* knockdown in S2 cells reproduced this phenotype (Fig. 2b), indicating that it is cell autonomous. Importantly, we grow S2 cells in serum-free medium, which lacks fatty acids normally bound to bovine serum albumin (BSA) in serum. Mitochondria of *Elov6*-knockdown cells rapidly re-fused upon addition of BSA-conjugated C18:0 to the medium for just 120 min (Fig. 2b) or 20 min (data not shown). Mitochondria of HeLa cells also fragmented when grown in medium with serum that was delipidated by organic extraction (Fig. 2c). This was rescued by re-adding BSA-conjugated C18:0 for 2 h (Fig. 2c), but not other fatty acids (Extended Data Fig. 2e). One lipid specific to mitochondria is lipoic acid. Knockdown of lipoic acid synthase (*LIAS*) led to reduced lipoic acid levels but not

mitochondrial fragmentation (Extended Data Fig. 2f), and depletion of C18:0 did not affect levels of lipoic acid or lipoylated proteins (Extended Data Fig. 2g, h), indicating that the effects of C18:0 are independent of lipoic acid. Thus, C18:0 regulates mitochondrial morphology in fly and human cells.

Mitochondrial fragmentation is due to either hyperactive fission or impaired fusion. Blocking fission with mdivi-1, a DRP1 inhibitor¹², induced mitochondrial fusion in control cells, but not in HeLa cells cultured without C18:0 (Fig. 2d), indicating impaired mitochondrial fusion in this condition. Mitochondria labelled with photoactivatable mitochondrially targeted green fluorescent protein (mitoGFP)¹³ rapidly fused with the rest of the network (stained with MitoTracker Red) in control cells, but not in cells cultured without C18:0, demonstrating impaired fusion (Fig. 2e).

Mitochondrial fusion is regulated largely by mitofusins^{14,15}. Epistasis experiments indicated that C18:0 acts upstream of mitofusin to regulate mitochondrial morphology: expression of *Drosophila* mitofusin (*dMfn*; also known as *Marf*) rescued mitochondrial fragmentation in *Elov6*-knockdown cells (Fig. 3a), indicating that *dMfn* acts downstream of C18:0. C18:0 did not induce fusion in the absence of *dMfn* (Fig. 3b), indicating that C18:0 requires *dMfn* for its action. Likewise, expression of *dMfn* rescued *Elov6*^{-/-} larval lethality (Fig. 3c). To test whether C18:0 can rescue *dMfn* loss of function, we generated *dMfn*-knockout flies. These flies phenocopy *Elov6*^{-/-} mutants: they have fragmented mitochondria, reduced mitochondrial respiration, and

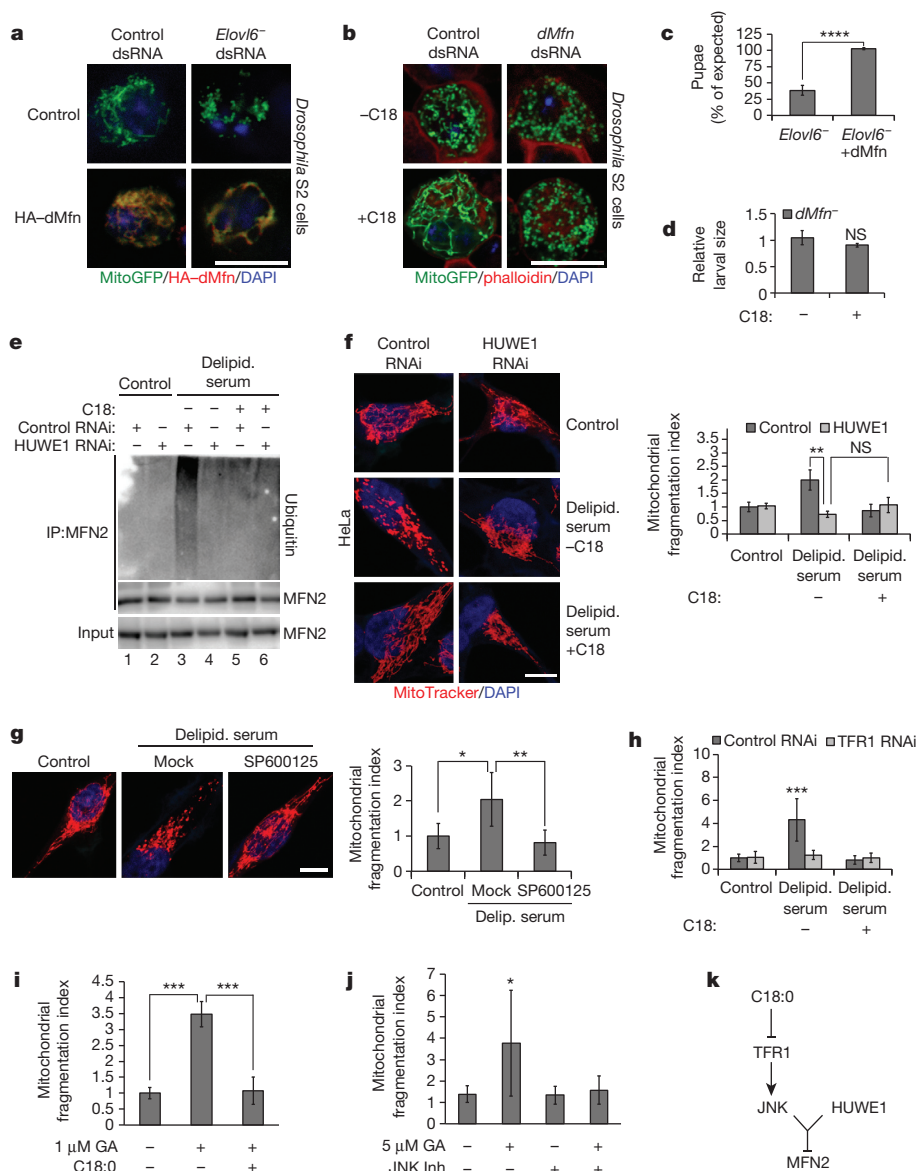


Figure 3 | C18:0 acts via TFR1, JNK and HUWE1 to regulate mitofusin. **a, b**, C18:0 acts upstream of *dMfn* to regulate mitochondrial morphology in S2 cells. Mitochondrial fragmentation induced by *Elovl6* knockdown is reversed by *dMfn* gain of function (**a**), whereas C18:0 (100 μ M, 2 h) cannot induce mitochondrial fusion in the absence of *dMfn* (**b**) ($n = 5$). dsRNA, double-stranded RNA. **c, d**, C18:0 acts upstream of *dMfn* to regulate *Drosophila* growth and survival. Ubiquitous expression of *dMfn* rescues lethality of *Elovl6* mutants until pupation (**c**) (χ^2 test, $P = 0.05$, $n = 685$), whereas C18:0 supplementation cannot rescue growth of *dMfn* mutant flies (**d**) ($n = 6$, not significant (NS) $P \geq 0.05$). **e, f**, Ubiquitination of endogenous MFN2 (**e**) and fragmentation of mitochondria (**f**) in response to C18:0 removal requires the Mfn2 ubiquitin ligase HUWE1. IP, immunoprecipitate. **g**, Pharmacological inhibition of JNK signalling (SP600125, 10 μ M) blunts mitochondrial fragmentation induced upon C18:0 removal (24 h delipidated serum). Representative images (left), quantification (right). $n = 15$. **h**, TFR1 is required for delipidated serum to induce mitochondrial fragmentation. For representative images, see Extended Data Fig. 7d. $n = 15$. **i**, Activation of TFR1 with 1 μ M gambogic acid (GA) leads to mitochondrial fragmentation, which is inhibited by 1 h C18:0 pre-treatment. For representative images, see Extended Data Fig. 9b. $n = 15$. **j**, JNK signalling is required to induce mitochondrial fragmentation in response to TFR1 activation with gambogic acid. HeLa cells treated with 10 μ M SP600125 before gambogic acid treatment (2 h, 1 μ M). Inh, inhibitor. For representative images, see Extended Data Fig. 9e. $n = 5$. **k**, Schematic diagram of the signalling route by which C18:0 regulates mitochondrial fusion. All scale bars, 10 μ m. Error bars show s.d. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, **** $P < 0.0001$, not significant (NS) $P \geq 0.05$, two-tailed t -test. For details, see Supplementary Methods.

die as early stage larvae that do not grow (Extended Data Fig. 3a–d). Dietary supplementation with C18:0 had no effect on the growth or viability of *dMfn* knockouts (Fig. 3d).

We asked whether C18:0 affects Mfn via post-translational modifications (PTMs). Mfn from *Elovl6*⁻ mutant larvae, or from HeLa cells growing with delipidated serum, migrated differently in SDS-polyacrylamide gel electrophoresis (SDS-PAGE) compared with control conditions (Extended Data Fig. 3e, f). Immunoprecipitating MFN2 from HeLa cells treated with or without C18:0 and probing with antibodies detecting various PTMs revealed that MFN2 from cells without C18:0 is hyper-ubiquitinated (Fig. 3e, lanes 1, 3, 5 (endogenous proteins) and Extended Data Fig. 3g (tagged proteins)). Several ubiquitin ligases target MFN2 (refs 16–18). Only knockdown of *HUWE1* rescued the mitochondrial fragmentation (Fig. 3f and Extended Data Fig. 4a–c) and MFN2 hyper-ubiquitination (Fig. 3e, lane 3 versus 4) caused by C18:0 removal, as well as lethality of *Elovl6* mutant flies (CG8184 in *Drosophila*; Extended Data Fig. 4d), identifying HUWE1 as the C18:0-responsive ubiquitin ligase. As expected¹⁹, increased MFN2 ubiquitination caused MFN2 protein destabilization (Extended Data Fig. 3h). C18:0 removal did not dramatically drop MFN2 steady-state levels, partly due to compensatory increases in MFN2 expression (Extended Data Fig. 3i), suggesting that

ubiquitination additionally blocks MFN2 function in a degradation-independent manner, as for other HUWE1 targets²⁰. HUWE1 only ubiquitinates MFN2 phosphorylated on Ser 27 by JNK¹⁹. Inhibition of JNK prevented mitochondrial fragmentation upon C18:0 removal (Fig. 3g). In sum, C18:0 regulates MFN2 ubiquitination via HUWE1, and thereby mitochondrial morphology and function. *Elovl6*⁻ mutant flies display other *dMfn* loss-of-function phenotypes, such as reduced endoplasmic-reticulum-mitochondrial connections and abnormal cristae^{21,22} (Extended Data Fig. 5).

We asked how C18:0 affects JNK or HUWE1 activity. Endoplasmic reticulum stress can activate JNK. However, C18:0 removal did not lead to an unfolded protein response (UPR) (Extended Data Fig. 6a, b) and neither knockdown of UPR effectors nor treatment with tauroursodeoxycholic acid (TUDCA), an endoplasmic reticulum chaperone that inhibits endoplasmic reticulum stress²³, blunted mitochondrial fragmentation upon C18:0 removal (Extended Data Fig. 6c, d). Instead, we hypothesized that C18:0 might regulate proteins via covalent binding ('stearylolation'), analogous to protein palmitoylation. We synthesized C18:0 derivatives with azide or alkyne functionalities, allowing covalent coupling to beads via copper-catalysed azide-alkyne cycloaddition ('click chemistry') (Extended Data Fig. 7b). We tested multiple derivatives, and only C17:0-azide induced mitochondrial

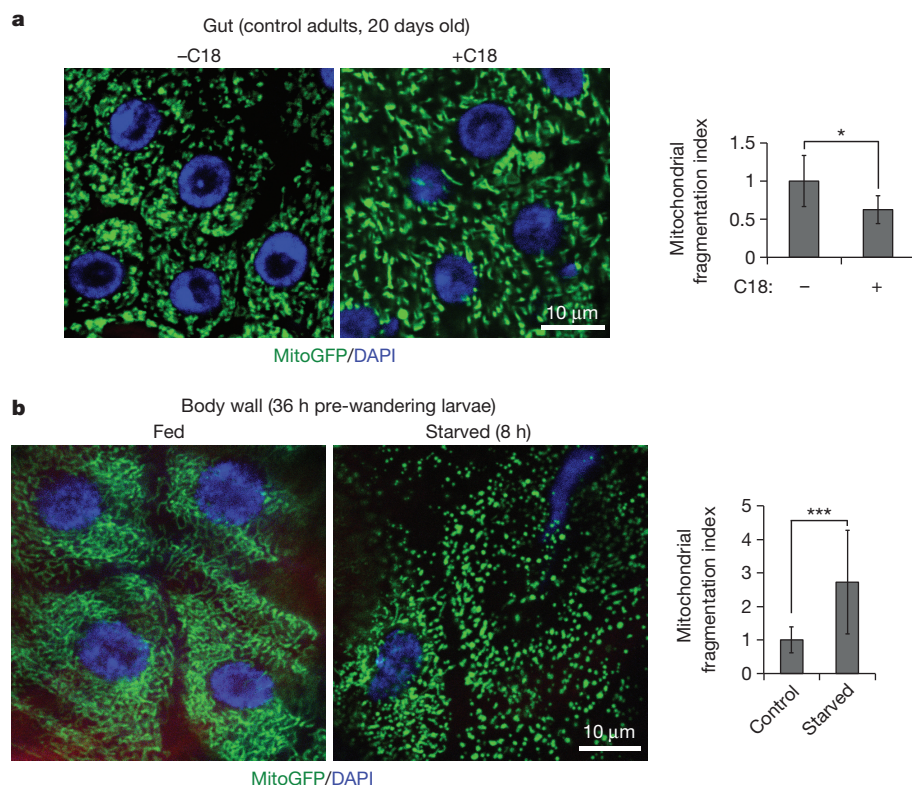


Figure 4 | Mitochondrial morphology is sensitive to dietary C18:0 levels in *Drosophila*. **a**, Dietary supplementation with 10% C18:0 leads to increased mitochondrial fusion in control flies (left), quantified as a drop in mitochondrial fragmentation (right) (4 animals per condition, 6 optical areas per animal). **b**, Starvation of larvae (8 h on PBS agar) causes mitochondrial fragmentation (left). Right, quantification (3 animals per condition, 16 optical areas per animal). For details, see Supplementary Methods. * $P < 0.05$, *** $P < 0.001$, two tailed t -test. Error bars represent s.d.

fusion like C18:0 (Extended Data Fig. 7a). We treated HeLa cells with C17:0-azide for 2 h, lysed them in 8 M urea to denature proteins, precipitated the lipid by coupling to beads, and identified covalently bound proteins by mass spectrometry. The most abundant protein in the lipid pull-down was TFR1 (Extended Data Fig. 7b, right), confirmed by immunoblotting (Extended Data Fig. 7c, lanes 1–3). Binding of TFR1 to C17:0-azide was abolished by treating lysates with hydroxylamine pH 7.5, indicating a thioester linkage (data not shown). We tested the top ten putatively stearoylated protein complexes for effects on mitochondrial morphology by knockdown. Knockdown of *TFR1* completely blunted mitochondrial fragmentation upon C18:0 removal (Fig. 3h and Extended Data Fig. 7b, right, d). Thus, TFR1 is required for cells to sense the absence of C18:0.

Since TFR1 is important for cellular iron uptake, TFR1 stearoylation could affect mitochondria via iron uptake or delivery. However, cells grow for days in medium lacking C18:0 but die in medium lacking iron (Extended Data Fig. 8a), suggesting that iron uptake is not markedly impaired in the absence of C18:0. Indeed, cells in medium lacking C18:0 do not show an iron deficiency transcriptional response (Extended Data Fig. 8b), a drop in protein or activity levels of enzymes containing iron-sulfur clusters (Extended Data Fig. 8c–f), impaired transferrin uptake (Extended Data Fig. 8g), or reduced association of transferrin with mitochondria (Extended Data Fig. 8h), suggesting that the effects of C18:0 are independent of iron. TFR1 also has a signalling function, activating JNK in response to the ligand gambogic acid²⁴. Low concentrations of gambogic acid that do not induce apoptosis (Extended Data Fig. 9a) induced rapid mitochondrial fragmentation in HeLa cells (2 h; Fig. 3i). This was suppressed by adding C18:0 (Fig. 3i and Extended Data Fig. 9b), indicating that C18:0 blocks this signalling function of TFR1. Indeed, treatment of HeLa cells with C18:0 reduced JNK activation, using JNK phosphorylation and phospho-JNK nuclear translocation as readouts (Extended Data Fig. 9c, d). JNK inhibition blocked the ability of gambogic acid to induce mitochondrial fragmentation (Fig. 3j and Extended Data Fig. 9e). In sum, these data suggest that TFR1 induces mitochondrial fragmentation via JNK, and that this is inhibited by TFR1 stearoylation (Fig. 3k).

Palmitoyl-transferases covalently bind C16:0 before transferring it to substrates. We found one member of this family, ZDHHC6, in our C17:0-azide pull-downs, suggesting that it is a C18:0 transferase. Indeed, knockdown of *ZDHHC6* blunted TFR1 stearoylation (Extended Data Fig. 7c, lane 5). Further work is required to study this in detail.

We noticed that elevating C18:0 levels in control cells increases mitochondrial fusion (Fig. 2b). Supplementing the diet of wild-type flies with C18:0 also increased mitochondrial fusion, whereas starvation of larvae led to mitochondrial fragmentation (Fig. 4). Thus, fly cells respond to both increases and decreases in levels of C18:0.

We asked whether dietary C18:0 supplementation could improve mitochondrial function in pathological conditions. Flies mutant for *Pink* or *Parkin* are established Parkinson's disease models. They have impaired mitochondrial function, and recapitulate Parkinson's disease phenotypes (reduced lifespan, neurodegeneration and impaired motor control^{25–29}). Dietary supplementation with C18:0 rescued the longevity, ATP levels and climbing defects of *Pink*[−] flies and the longevity of *Parkin*[−] flies (Fig. 4 and Extended Data Fig. 10; other *Parkin* phenotypes not tested).

We identify C18:0 as a regulator of mitochondrial function. Upon loss of C18:0, TFR1 de-stearoylation activates JNK, leading to HUWE1-dependent MFN2 ubiquitination, impaired MFN2 activity¹⁹, and mitochondrial fragmentation. Loss of C18:0 in flies specifically impacts mitochondrial function, since *Elovl6*[−] lethality can be rescued by *Spargel*, *AOX* or *dMfn* expression or *Huwei1* knockdown. To our knowledge, this is the first time stearoylation of a human protein has been found to regulate its function. The link between TFR1 and mitochondria perhaps makes sense, because iron enters cells via TFR1 and then mainly travels to mitochondria for iron-sulfur clusters. Flies are sensitive to dietary C18:0; increased dietary C18:0 leads to increased mitochondrial fusion *in vivo*. Thus, the metabolite C18:0 acts as a signalling molecule linking diet to mitochondrial function. Intriguingly, dietary C18:0 can also improve mitochondrial function in some pathological conditions in the fly, since dietary supplementation with C18:0 improved the Parkinson's-disease-related

phenotypes observed in *Pink* and *Parkin* mutant flies (see Supplementary Discussion).

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 17 April 2014; accepted 1 June 2015.

Published online 27 July; corrected online 2 September 2015 (see full-text HTML version for details).

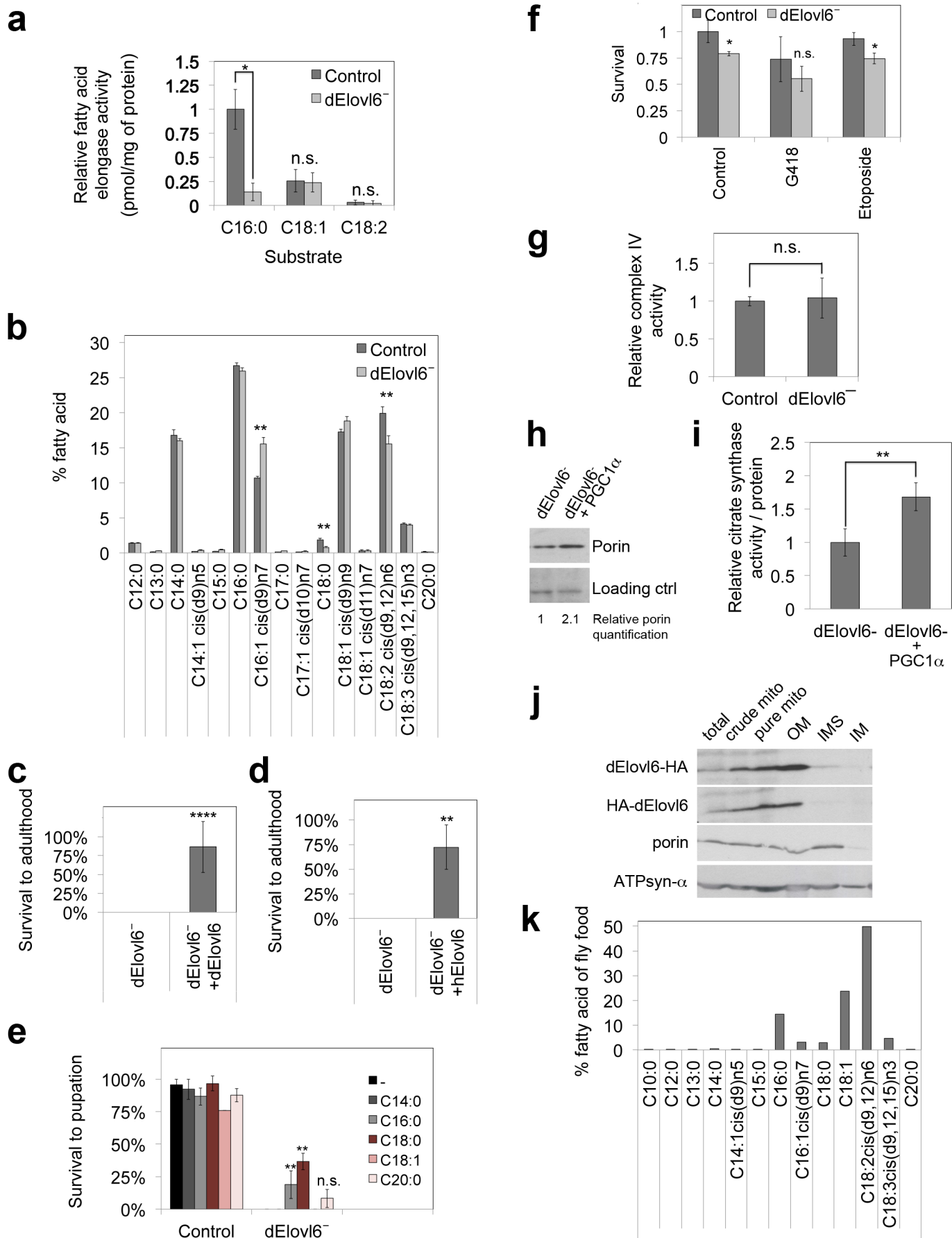
- McBride, H. M., Neuspiel, M. & Wasiak, S. Mitochondria: more than just a powerhouse. *Curr. Biol.* **16**, R551–R560 (2006).
- Vafai, S. B. & Mootha, V. K. Mitochondrial disorders as windows into an ancient organelle. *Nature* **491**, 374–383 (2012).
- Suen, D. F., Norris, K. L. & Youle, R. J. Mitochondrial dynamics and apoptosis. *Genes Dev.* **22**, 1577–1590 (2008).
- Nunnari, J. & Suomalainen, A. Mitochondria: in sickness and in health. *Cell* **148**, 1145–1159 (2012).
- Jornayvaz, F. R. & Shulman, G. I. Regulation of mitochondrial biogenesis. *Essays Biochem.* **47**, 69–84 (2010).
- Moon, Y. A., Ochoa, C. R., Mitsche, M. A., Hammer, R. E. & Horton, J. D. Deletion of ELOVL6 blocks the synthesis of oleic acid but does not prevent the development of fatty liver or insulin resistance. *J. Lipid Res.* **55**, 2597–2605 (2014).
- Matsuzaka, T. *et al.* Elov6 promotes nonalcoholic steatohepatitis. *Hepatology* **56**, 2199–2208 (2012).
- Jung, A., Hollmann, M. & Schafer, M. A. The fatty acid elongase NOA is necessary for viability and has a somatic role in *Drosophila* sperm development. *J. Cell Sci.* **120**, 2924–2934 (2007).
- Fernandez-Ayala, D. J. *et al.* Expression of the *Ciona intestinalis* alternative oxidase (AOX) in *Drosophila* complements defects in mitochondrial oxidative phosphorylation. *Cell Metab.* **9**, 449–460 (2009).
- Friedman, J. R. & Nunnari, J. Mitochondrial form and function. *Nature* **505**, 335–343 (2014).
- Okamoto, K. & Shaw, J. M. Mitochondrial morphology and dynamics in yeast and multicellular eukaryotes. *Annu. Rev. Genet.* **39**, 503–536 (2005).
- Cassidy-Stone, A. *et al.* Chemical inhibition of the mitochondrial division dynamin reveals its role in Bax/Bak-dependent mitochondrial outer membrane permeabilization. *Dev. Cell* **14**, 193–204 (2008).
- Twig, G. *et al.* Tagging and tracking individual networks within a complex mitochondrial web with photoactivatable GFP. *Am. J. Physiol. Cell Physiol.* **291**, C176–C184 (2006).
- Campello, S. & Scorrano, L. Mitochondrial shape changes: orchestrating cell pathophysiology. *EMBO Rep.* **11**, 678–684 (2010).
- Liesa, M., Palacin, M. & Zorzano, A. Mitochondrial dynamics in mammalian health and disease. *Physiol. Rev.* **89**, 799–845 (2009).
- Escobar-Henriques, M. & Langer, T. Dynamic survey of mitochondria by ubiquitin. *EMBO Rep.* **15**, 231–243 (2014).
- Livnat-Levanon, N. & Glickman, M. H. Ubiquitin–proteasome system and mitochondria—reciprocity. *Biochim. Biophys. Acta* **1809**, 80–87 (2011).
- Ziviani, E., Tao, R. N. & Whitworth, A. J. *Drosophila* Parkin requires PINK1 for mitochondrial translocation and ubiquitinates Mitofusin. *Proc. Natl Acad. Sci. USA* **107**, 5018–5023 (2010).
- Leboucher, G. P. *et al.* Stress-induced phosphorylation and proteasomal degradation of mitofusin 2 facilitates mitochondrial fragmentation and apoptosis. *Mol. Cell* **47**, 547–557 (2012).
- de Groot, R. E. *et al.* Huwe1-mediated ubiquitylation of Dishevelled defines a negative feedback loop in the Wnt signaling pathway. *Sci. Signal.* **7**, ra26 (2014).
- Lee, S. *et al.* Mitofusin 2 is necessary for striatal axonal projections of midbrain dopamine neurons. *Hum. Mol. Genet.* **21**, 4827–4835 (2012).
- de Brito, O. M. & Scorrano, L. Mitofusin 2 tethers endoplasmic reticulum to mitochondria. *Nature* **456**, 605–610 (2008).
- Ozcan, U. *et al.* Chemical chaperones reduce ER stress and restore glucose homeostasis in a mouse model of type 2 diabetes. *Science* **313**, 1137–1140 (2006).
- Jian, J., Yang, Q. & Huang, X. Src regulates Tyr²⁰ phosphorylation of transferrin receptor-1 and potentiates breast cancer cell survival. *J. Biol. Chem.* **286**, 35708–35715 (2011).
- Whitworth, A. J., Wes, P. D. & Pallanck, L. J. *Drosophila* models pioneer a new approach to drug discovery for Parkinson's disease. *Drug Discov. Today* **11**, 119–126 (2006).
- Greene, J. C. *et al.* Mitochondrial pathology and apoptotic muscle degeneration in *Drosophila parkin* mutants. *Proc. Natl Acad. Sci. USA* **100**, 4078–4083 (2003).
- Park, J. *et al.* Mitochondrial dysfunction in *Drosophila PINK1* mutants is complemented by *parkin*. *Nature* **441**, 1157–1161 (2006).
- Park, J., Kim, Y. & Chung, J. Mitochondrial dysfunction and Parkinson's disease genes: insights from *Drosophila*. *Dis. Model. Mech.* **2**, 336–340 (2009).
- Guo, M. *Drosophila* as a model to study mitochondrial dysfunction in Parkinson's disease. *Cold Spring Harb. Perspect. Med.* **2**, a009944 (2012).
- Jump, D. B. Mammalian fatty acid elongases. *Methods Mol. Biol.* **579**, 375–389 (2009).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank A. Zorzano and J. Seco for photo-activatable mitoGFP, A. Zorzano and J. Pablot for Myc-tagged MFN2, H. Jacobs for UAS-AOX flies, C. Frei for UAS-Spargel flies, P. Verstreken for Parkinson's flies, A. Whitworth for anti-dMfn, J. Chung for anti-dPorin, R. Lill for anti-IBA57, J. Sagalés for help with respirometry on the *dMfn*-knockout animals, F. Melchior for advice on detecting ubiquitination, K. Richter and the DKFZ EM Core Facility for help with electron microscopy, the DKFZ Light Microscopy Core Facility for help with live-cell imaging, M. Schnölzer and the DKFZ Proteomics Core Facility for mass spectrometry analysis of proteins, and M. del Sol for technical support. This work was supported in part by Deutsche Forschungsgemeinschaft (DFG), SFB1118, Helmholtz Portfolio Topic 'Metabolic Dysfunction', and European Research Council Starting Grants (#260602) to A.A.T., and Biotechnology and Biological Sciences Research Council, Medical Research Council programme and British Heart Foundation grants to A.V.-P.

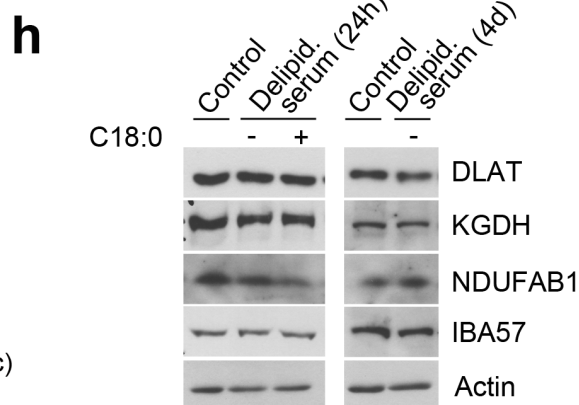
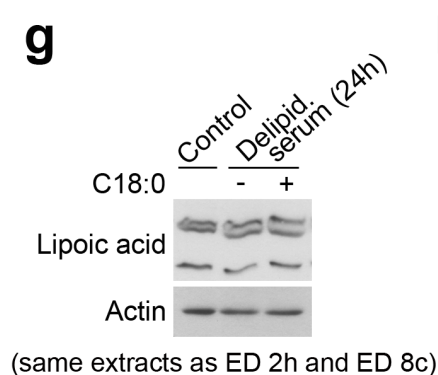
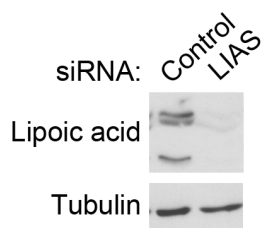
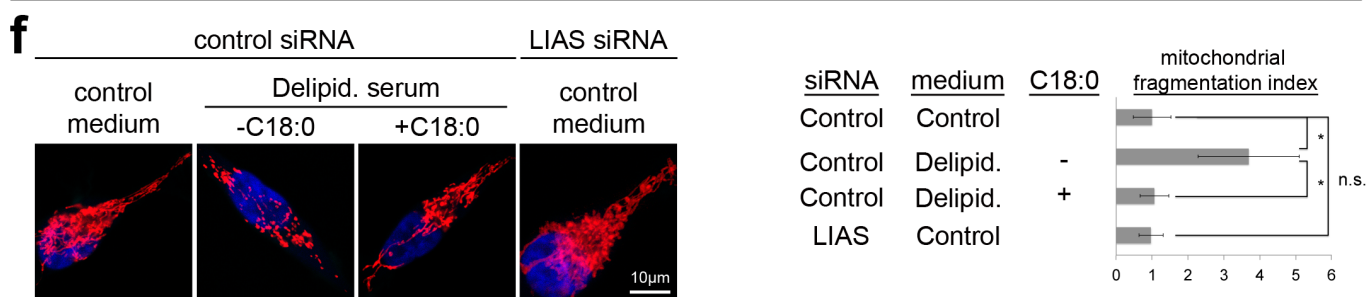
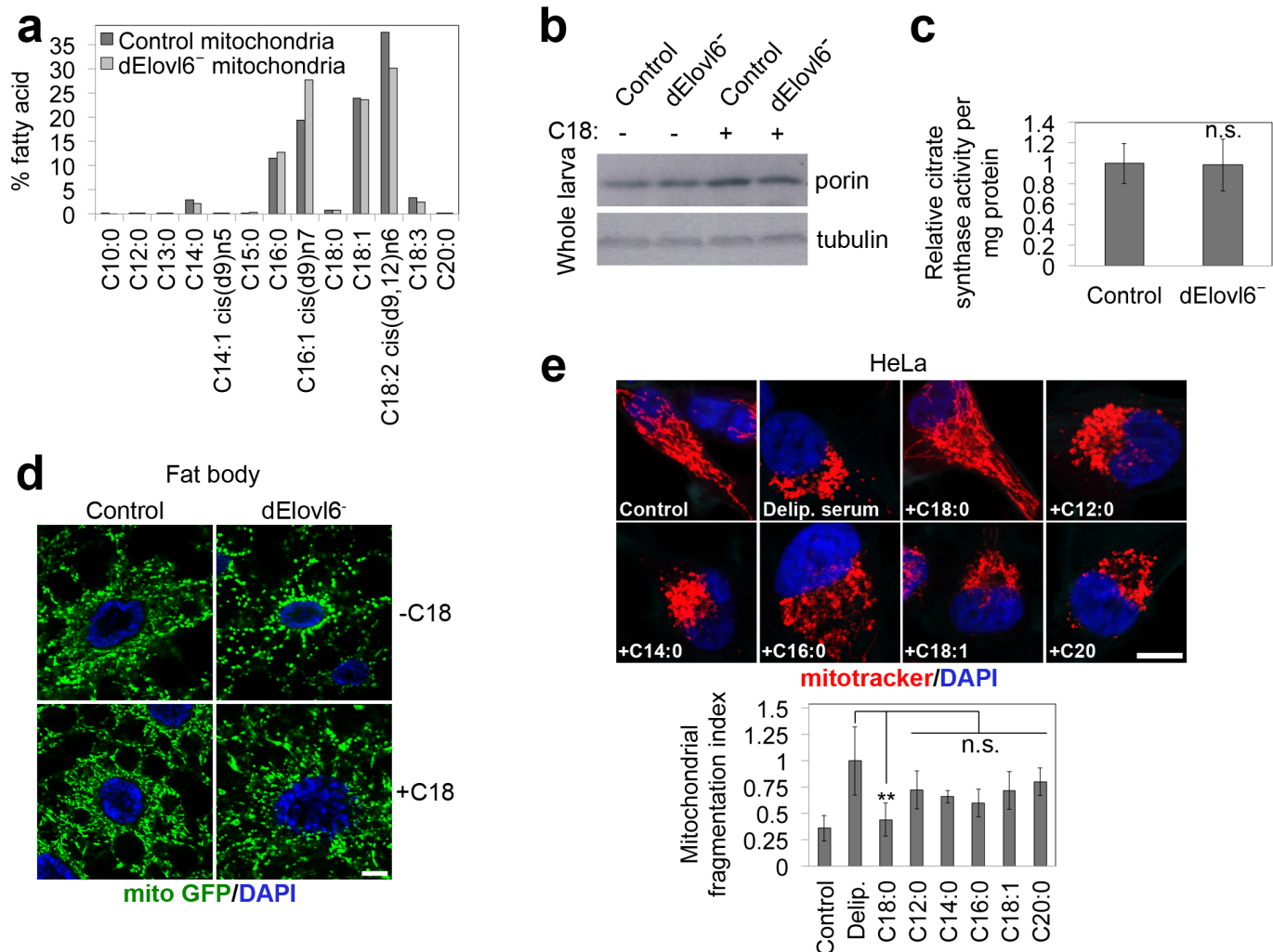
Author Contributions All authors designed experiments. D.S., S.V., X.X. and C.Y.T. performed experiments. All authors analysed data. D.S., S.V., A.K.M., A.V.-P. and A.A.T. wrote the manuscript and all authors read and edited the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.A.T. (a.teleman@dkfz.de).



Extended Data Figure 1 | noa/Elovl6 is the functional homologue of human ELOVL6. **a**, C16:0 to C18:0 elongase activity is significantly blunted in *Elovl6* mutants, whereas elongase activities on other fatty acids measured are not affected. Microsomal preparations from control or *Elovl6* mutant animals were incubated with radioactive malonyl-CoA and the indicated fatty-acyl-CoA. Elongation was quantified by incorporation of the aqueous metabolite malonyl-CoA into lipid-soluble fatty acids, as described previously³⁰. Values represent biological triplicates. **b**, Gas chromatography flame ionization detector (GC-FID) analysis reveals that *Elovl6* mutant larvae have reduced levels of C18:0, the product of *Elovl6* elongase activity. Values are the averages of technical duplicates on biological duplicates. Error bars represent standard error of the mean (s.e.m.). **c**, Lethality of *Elovl6* mutants is fully rescued to expected Mendelian ratios by ubiquitous expression (with actin-GAL4) of *Elovl6* from a *UAS* transgene (χ^2 test, $1.149 < 3.841 = \chi^2$, where $P = 0.05$, $n = 141$; **** $P < 0.0001$). **d**, Human ELOVL6 and *Drosophila* Elovl6 are functionally equivalent, since the lethality of *Elovl6* mutant flies is fully rescued to expected Mendelian ratios by ubiquitous expression (with actin-GAL4) of human ELOVL6 from a *UAS* transgene (χ^2 test, $2.38 < 3.841 = \chi^2$, where $P = 0.05$, $n = 76$; ** $P < 0.01$). **e**, The lethality of *Elovl6* mutant flies is most strongly rescued by C18:0, the product of *Elovl6*. Synchronized 1st instar larvae of indicated genotypes were grown on standard food supplemented with indicated fatty acids (5%). The percentage of total animals surviving to pupation was calculated. Values represent average of biological triplicates. **f**, *Elovl6* mutants are not hypersensitive to drugs such as G418 (protein biosynthesis inhibitor) or etoposide (topoisomerase inhibitor). Thirty synchronized L1 larvae were grown in vials with food supplemented with either G418 (50 $\mu\text{g ml}^{-1}$) or etoposide (25 μM). Percentage of animals that reach

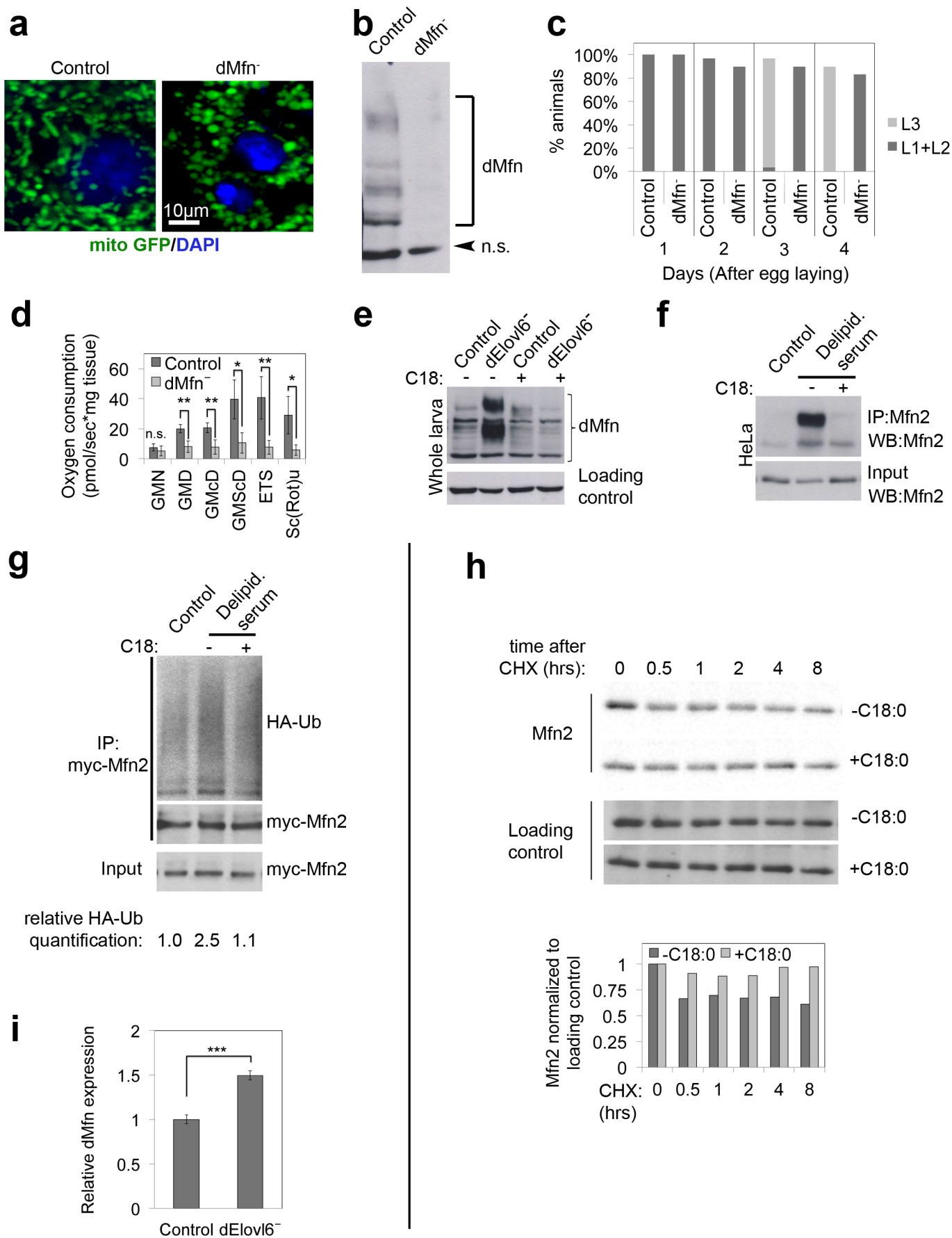
pupation was quantified. Values represent average of four biological replicates. **g**, Complex IV activity of *Elovl6*⁻ larvae is not impaired. Complex IV activity of female pre-wandering larvae was measured with Oroboros high-resolution respirometry. Oxygen consumption was measured in the presence of only *N,N,N',N'*-tetramethyl-*p*-phenylenediamine dihydrochloride (TMPD) as substrate, which can be directly oxidized by complex IV. The values were corrected for non-mitochondrial oxygen consumption (oxygen consumption in the presence of complex IV inhibitor potassium cyanide (KCN)) and normalized to tissue weight. $n = 3$. **h, i**, Overexpression of *Spargel* in *Elovl6* mutant female pre-wandering larvae leads to increased mitochondrial abundance, assessed by porin levels (**h**; representative of six biological replicates) and citrate synthase activity (**i**; $n = 4$) in pre-wandering larvae. See Supplementary Fig. 12 for image of the uncropped full western blot. **j**, *Drosophila* Elovl6 (either N- or C-terminally tagged) localizes to the mitochondrial outer membrane. S2 cell lysates ('total') were successively fractionated to yield crude mitochondria (which include mitochondrial-associated membranes (MAMs)), pure mitochondria (lacking MAMs), and mitochondrial outer membranes (OM), inner membranes (IM) and inter-membrane space (IMS). Endogenous porin and ATPsyn- α were used as positive controls for outer membranes and inner membranes, respectively. 7.5 μg of protein from each fraction was loaded per lane. See Supplementary Fig. 13 for image of the uncropped full western blot. Representative of two biological replicates. **k**, Lipidomic analysis of standard fly food reveals low levels of C18:0 in the food. **a, c, d, e, f, g, i**, Error bars represent s.d. **a, b, e-g, i**, * $P < 0.05$, ** $P < 0.01$, not significant (NS) $P \geq 0.05$, two-tailed *t*-test.



Extended Data Figure 2 | C18:0 regulates mitochondrial morphology.

a, Lipidomic (GC-FID) profiles of purified mitochondria from *Elovl6* mutant 3rd instar larvae do not show major differences compared with control animals. Mitochondrial membranes from both control and mutant animals have very low levels of C18:0. Controls for purity of mitochondrial prep are shown in Extended Data Fig. 6b. **b, c**, *Elovl6* mutant larvae do not have reduced amounts of mitochondria, quantified via levels of porin (**b**; representative of three biological replicates) or citrate synthase activity (**c**; $n = 3$). See Supplementary Fig. 14 for image of the uncropped full western blot. **d**, *Elovl6* mutant larvae have fragmented mitochondria, which is rescued by dietary C18:0 supplementation. Mitochondrial morphology from fat bodies of control or *Elovl6* mutant female larvae, fed control or C18:0 (10%) supplemented food, visualized with mitoGFP. Images are representative of eight areas of four larvae from each genotype and food conditions. Equivalent pictures for body wall are shown in Fig. 2a. **e**, Only C18:0, and not shorter, longer or desaturated fatty acids, restores mitochondrial fragmentation to control levels in HeLa cells grown in medium containing delipidated serum. Mitochondria were visualized with MitoTracker (red) (top) and mitochondrial fragmentation was quantified by normalizing the number of mitochondrial particles to total

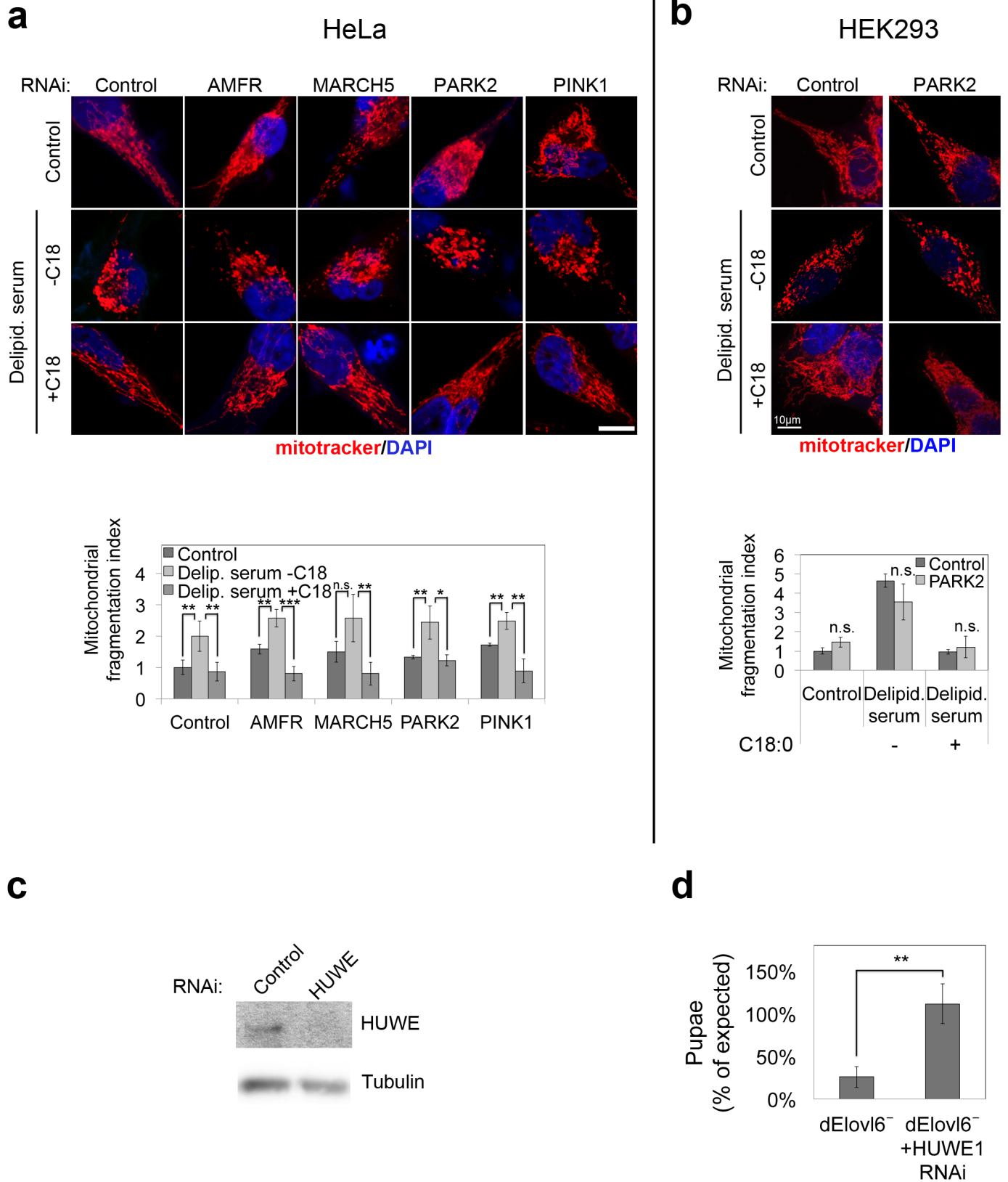
mitochondrial area (bottom) ($n = 15$). **f**, Reduced lipoic acid (LA) levels do not lead to mitochondrial fragmentation. Lipoic acid synthase (*LIAS*) was knocked down by RNAi in HeLa cells, leading to significantly reduced lipoic acid levels, assayed by immunoblotting of total cell lysates with antibody detecting lipoic acid (bottom left). Unlike removal of C18:0, this does not lead to mitochondrial fragmentation. Representative images (top left) are quantified (top right) ($n = 6$). See Supplementary Fig. 14 for image of the uncropped full western blot. **g, h**, HeLa cells growing in medium containing delipidated serum do not display reduced levels of protein lipoylation (**g**) or reduced levels of lipoylated proteins (**h**). HeLa cells were grown in medium containing delipidated serum for either 24 h (the same time point used for all other experiments in which mitochondrial fragmentation was assessed) (**g, h**), or for an extended period of time: 4 days (**h**). Lipoic acid levels were assayed by immunoblotting total cell lysates with an anti-lipoic acid antibody (**g**), and levels of lipoylated proteins were assessed with specific antibodies (**h**). See Supplementary Fig. 15 for image of the uncropped full western blot. **c, e, f**, $*P < 0.05$, $**P < 0.01$, not significant (NS) $P \geq 0.05$, two-tailed t -test. Error bars represent s.d. Scale bars, 10 μm .



Extended Data Figure 3 | Mitofusin loss of function phenocopies *Elovl6* mutation or removal of C18:0.

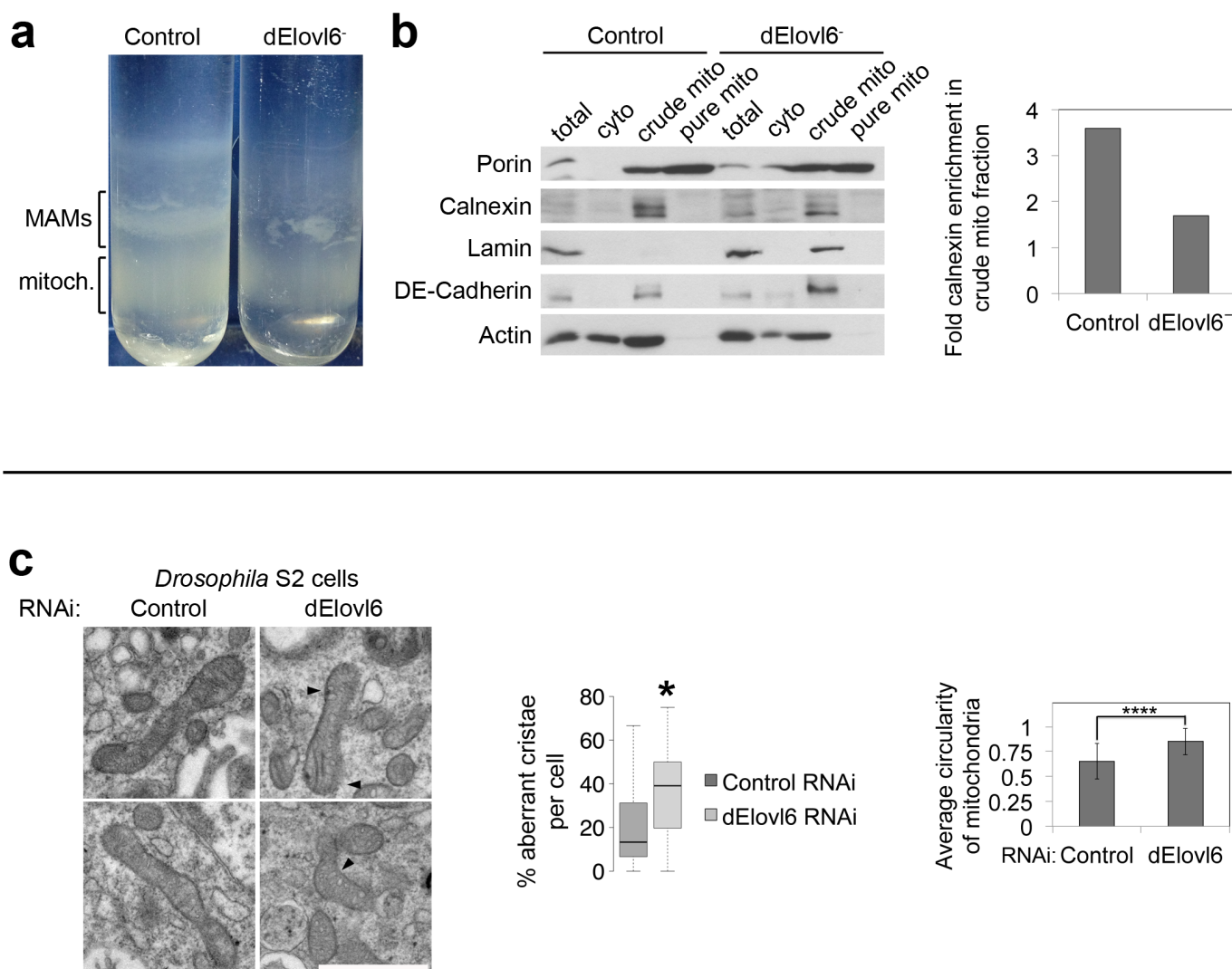
a, *dMfn* knockout larvae (1st instar) have fragmented mitochondria, visualized with mitoGFP. Representative of ten images. **b**, Endogenous dMfn runs as a main band plus a laddering of apparently increasing molecular weights on an SDS-PAGE gel. Specificity is controlled by blotting lysates from control and *Mfn*-knockout female larvae with anti-Mfn antibody. **c**, Homozygous mutation of *dMfn* is lethal. *Mfn*-knockout larvae survive for several days as small L1/L2 larvae and eventually die. Synchronized 1st instar larvae were grown on standard fly food and examined every 24 h for developmental stage and per cent survival ($n = 30$). **d**, *dMfn*-knockout animals have impaired oxygen consumption. Oxygen consumption of inverted, digitonin permeabilized, female larval tissues was measured with an Oroboros oxygraph chamber and normalized to tissue weight. Oxygen consumption was measured in the presence of the following substrates: GMN (glutamate and malate), GMD (glutamate, malate and ADP), GMCD (glutamate, malate, cytochrome *c* and ADP), GMScD (glutamate, malate, succinate, cytochrome *c* and ADP), ETS (glutamate, malate, cytochrome *c*, ADP and uncoupling reagent), and Sc(Rot)u (glutamate, malate, cytochrome *c*, ADP and rotenone). $n = 5$. **e**, Endogenous dMfn is post-translationally modified in a C18:0-dependent manner in *Drosophila*. dMfn from *Elovl6* female mutants migrates in an SDS-PAGE gel differently, compared with Mfn2 from control animals. This is reversed by supplementing the diet with C18:0. All indicated bands are dMfn, since they disappear in lysates from *dMfn*-knockout animals (see Extended Data Fig. 3b). Flies were grown on antifungal-free food. **f**, Endogenous MFN2 is post-translationally

modified in a C18:0-dependent manner in human HeLa cells. MFN2 immunoprecipitated from HeLa cells treated for 24 h with medium containing standard or delipidated serum, and then for 2 h in the absence or presence of C18:0 (100 μ M), lysed in 8M urea (see Methods). **g**, C18:0 affects ubiquitination of MFN2. MFN2 is more heavily ubiquitinated in cells treated with delipidated serum than in control cells and this is reversed by supplementing the medium with C18. HeLa cells were cotransfected with tagged versions of MFN2 (*myc*) and ubiquitin (*HA*). Tagged MFN2 was immunoprecipitated and blots were probed with HA antibody to detect ubiquitination. Quantification of ubiquitination, normalized to Myc-MFN2 in the immunoprecipitate (IP) is shown below the lane. **h**, C18:0 removal destabilizes MFN2 protein. A cyclohexamide (CHX) chase experiment was performed to block *de novo* synthesis of MFN2, thereby looking at turnover of existing MFN2 protein *in vivo*. HeLa cells treated with medium containing delipidated serum plus or minus C18:0 were treated with 100 μ M CHX and then lysed at the indicated time points to compare MFN2 protein levels. Bottom, densitometric quantification of the blots normalized to loading control. **i**, *dMfn* expression is upregulated in *Elovl6*⁻ flies compared with controls. *dMfn* transcript levels in 24 h female pre-wandering larvae were determined by quantitative polymerase chain reaction with reverse transcription (RT-PCR), normalized to rp49 (in triplicates). Scale bar, 10 μ m. **d, i**, * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, not significant (NS) $P \geq 0.05$, two-tailed *t*-test. Error bars represent s.d. See Supplementary Fig. 16 for images of the uncropped full western blots.



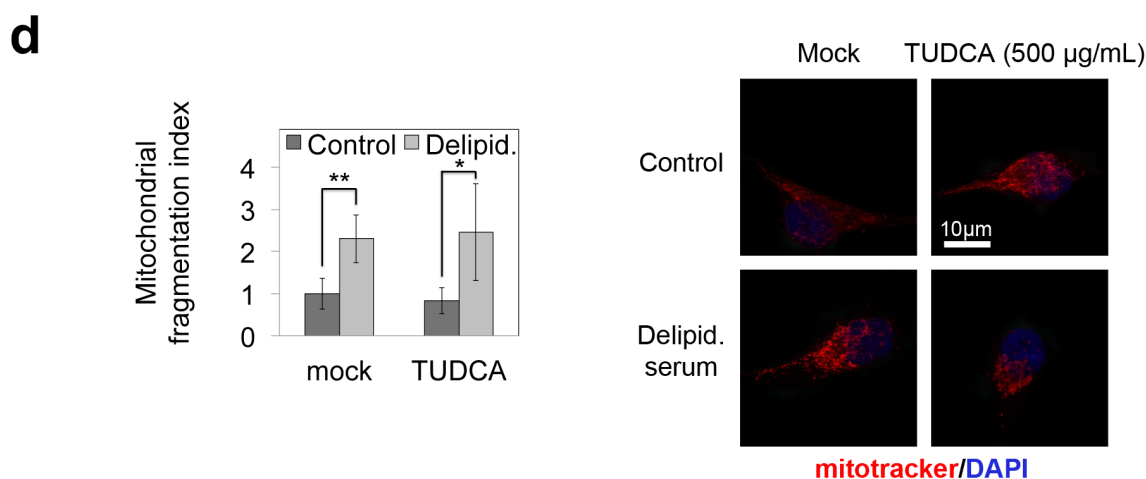
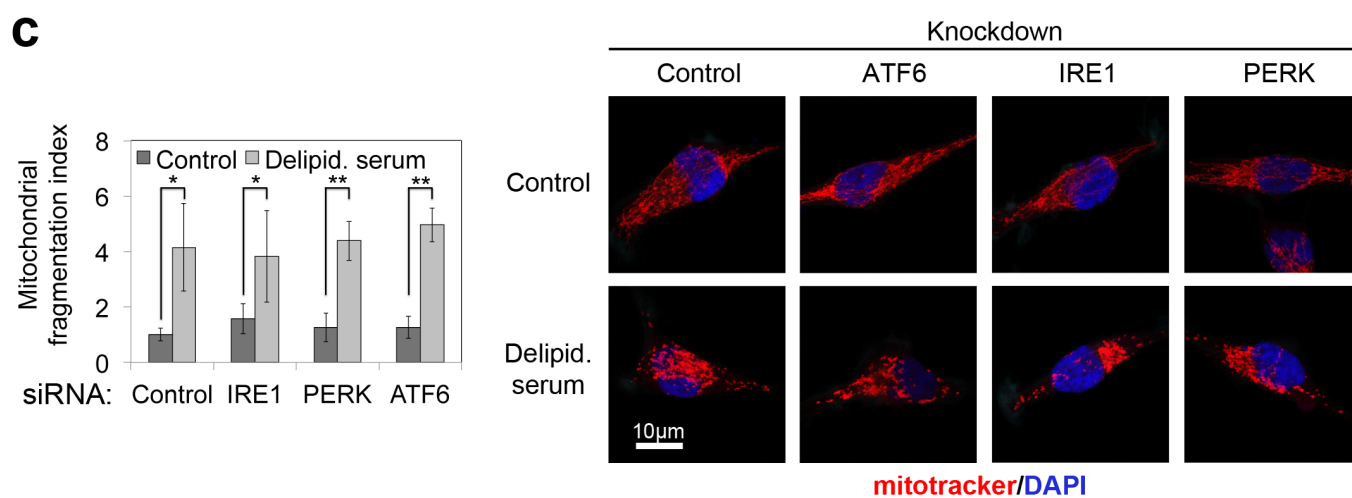
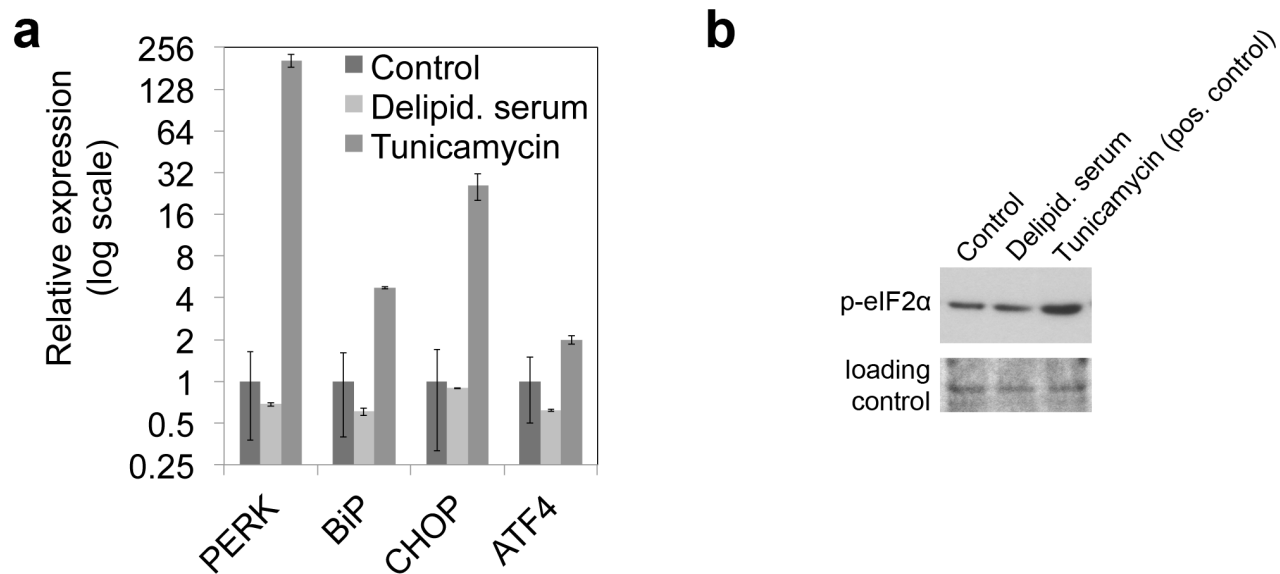
Extended Data Figure 4 | HUWE1 is required for hyperubiquitination of MFN2 in response to C18:0 withdrawal. **a**, Short interfering RNA (siRNA) depletion of other ubiquitin ligases targeting MFN (besides HUWE1, shown in Fig. 3) does not rescue the mitochondrial fragmentation induced by removal of C18:0 (top). Bottom, quantification. $n = 15$. **b**, siRNA depletion of *PARK2* in HEK293 cells, as in HeLa cells (**a**), does not rescue the mitochondrial fragmentation induced by removal of C18:0 (top). Bottom, quantification. $n = 15$. **c**, *HUWE1* knockdown efficiency controlled by detecting HUWE1 protein levels. See Supplementary Fig. 17 for image of the uncropped full

western blot. **d**, Survival to pupation of *Elovl6* mutants is fully rescued by ubiquitous expression (*daughterless-GAL4*) of RNAi targeting *Huwe1* (*CG8184*). *Elovl6* mutants expressing *HUWE1* RNAi survive to pupation at expected Mendelian frequencies (χ^2 test, $0.86 < 3.841 = \chi^2$, where $P = 0.05$). Flies were grown on antifungal-free food. Values represent average of four biological replicates. Scale bars, 10 μm . **a**, **b**, **d**, $*P < 0.05$, $**P < 0.01$, $***P < 0.001$, not significant (NS) $P \geq 0.05$, two-tailed t -test. Error bars represent s.d.



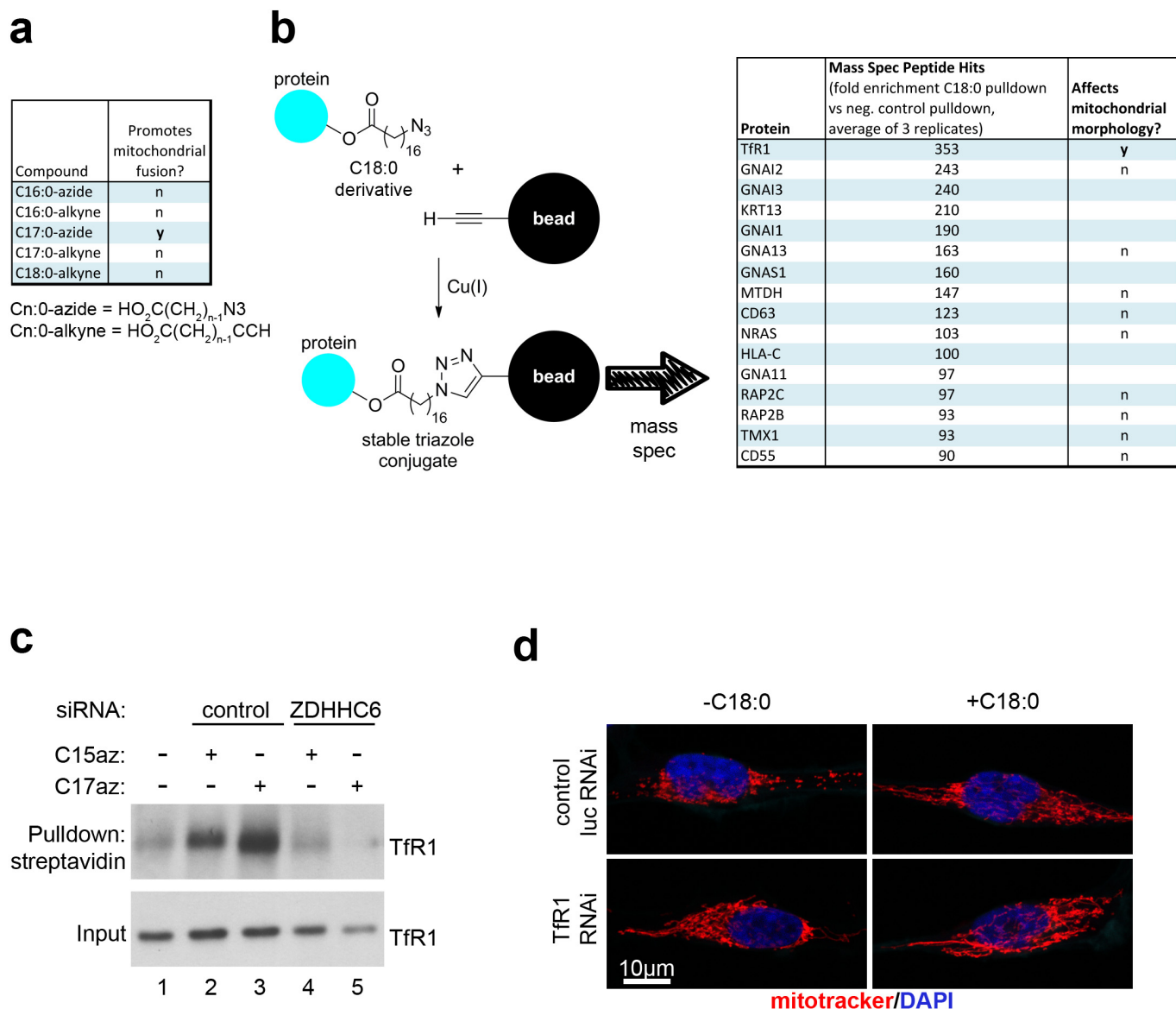
Extended Data Figure 5 | *Elov6* mutants have other Mfn loss-of-function phenotypes such as reduced mitochondrial-associated membranes and abnormal cristae. **a**, The mitochondrial associated membranes (MAM) band is strongly reduced or absent in Percoll gradients of crude mitochondrial fractions from *Elov6*-mutant animals, compared with controls. **b**, Purity control of mitochondrial preparations show that pure mitochondrial fractions are lacking markers of other subcellular organelles such as calnexin (endoplasmic reticulum) and lamin (nuclei). See Supplementary Fig. 17 for image of the uncropped full western blot. Right, quantification shows that levels of the endoplasmic reticulum marker calnexin are reduced in crude

mitochondrial fractions from *Elov6* mutants, compared with controls, in agreement with reduced MAMs in *Elov6* mutants. Values show densitometry ratios of calnexin levels in crude mitochondrial fractions, normalized to total lysate calnexin. **c**, Electron microscopy of *Drosophila* S2 cell mitochondria (left) reveals cristae abnormalities in *Elov6*-depleted cells. Middle, quantification ($n = 200$). Significance of the difference was calculated with a Mann–Whitney test ($*P < 0.05$). Right, average circularity of mitochondria was calculated with ImageJ software. Scale bar, 1 μm . $n = 200$, $****P < 0.0001$, two-tailed t -test. Error bars show s.d.



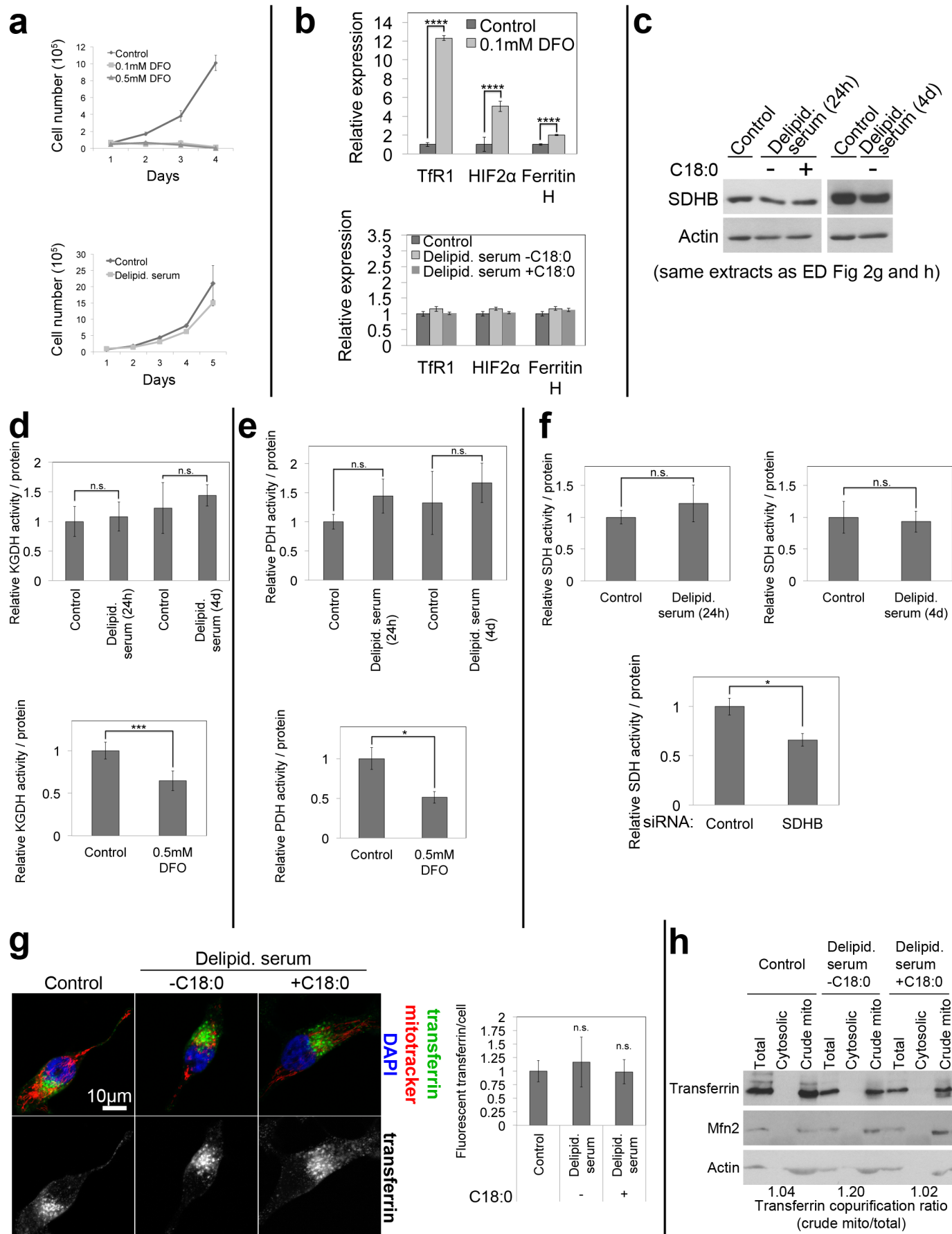
Extended Data Figure 6 | C18:0 removal does not lead to endoplasmic reticulum stress, and inhibiting UPR does not inhibit mitochondrial fragmentation upon C18:0 removal. **a**, C18:0 removal for 24 h does not induce expression of UPR target genes, quantified by quantitative RT-PCR, normalized to *RPL19*. *BiP* (also known as *HSPA5*) is a readout for IRE1 (also known as ERN1) activation, *CHOP* (also known as *DDIT3*) is a readout for ATF6 activation, and *PERK* (also known as *EIF2AK3*) is a readout of its own activation due to a positive transcriptional feedback loop. Tunicamycin serves as a positive control. *y* axis is displayed in a logarithmic scale to fit all data points on one graph. The experiment was done in triplicates. **b**, p-eIF2 α , a UPR marker, does not increase upon removal of C18:0 whereas it is induced

by tunicamycin, a positive control. See Supplementary Fig. 18 for image of the uncropped full western blot. **c**, Knocking down mediators of the UPR response does not inhibit mitochondrial fragmentation upon C18:0 removal. HeLa cells were transfected with either control siRNAs or siRNAs targeting UPR mediators as indicated. Left, the mitochondrial fragmentation index; right, representative images. *n* = 15. **d**, Inhibiting endoplasmic reticulum stress by means of a chemical chaperone, TUDCA, does not rescue mitochondrial fragmentation upon C18:0 removal. HeLa cells were pre-treated with 500 $\mu\text{g ml}^{-1}$ TUDCA 30 min before delipidated serum treatment. Left, mitochondrial fragmentation index (*n* = 15); right, representative images. **a**, **c**, **d**, **P* < 0.05, ***P* < 0.01, two-tailed *t*-test. Error bars show s.d.



Extended Data Figure 7 | TFR1 is the mediator of C18:0 signalling to mitochondrial morphology. **a**, C17:0-azide is a functional analogue of C18:0 in that it induces mitochondrial fusion in HeLa cells, whereas other C18:0 derivatives are not. Cn:0-azide = $\text{HO}_2\text{C}(\text{CH}_2)_{n-1}\text{N}_3$; Cn:0-alkyne = $\text{HO}_2\text{C}(\text{CH}_2)_{n-1}\text{CCH}$. **b**, TFR1 is the most enriched protein in a C17:0-azide pulldown, and it regulates mitochondrial morphology. HeLa cells were treated with C17:0-azide for 2 h, and covalently bound proteins were precipitated by lysing cells under denaturing conditions (8 M urea), and linking the C17:0-azide to an alkyne-labelled resin via click chemistry (left). Precipitated proteins were identified by mass spectrometry, and peptide counts were normalized to peptide counts in a negative control pulldown from cells not treated with C17:0-azide ($n = 3$) (right; column 2). Indicated proteins were also tested by siRNA-mediated knockdown for effects on mitochondrial morphology (column 3). **c**, TFR1 is covalently bound to the C18:0 derivative C17:0-azide in HeLa cells in a ZDHHC6-dependent manner. HeLa cells were treated with

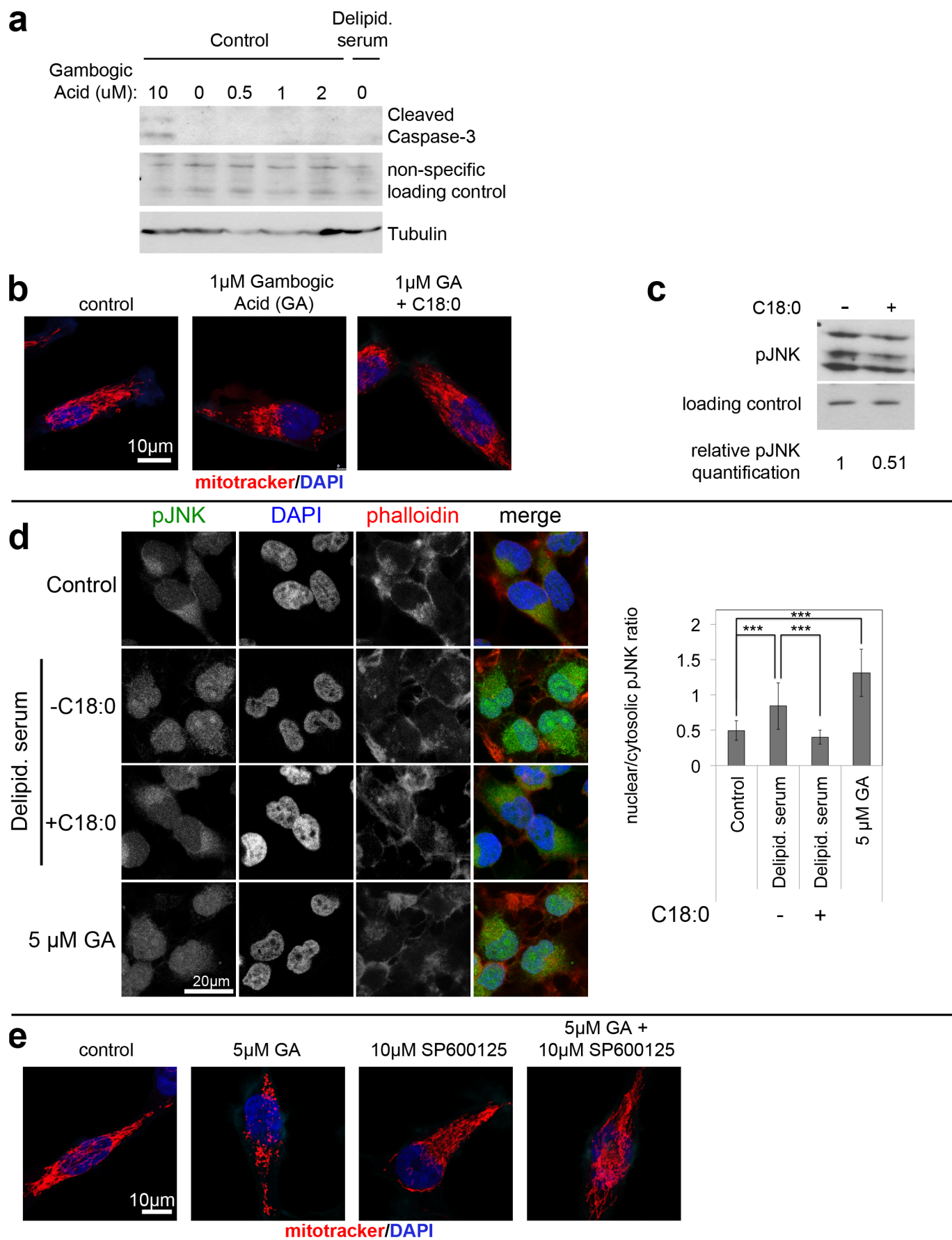
C17:0-azide for 2 h, and subsequently lysed in denaturing conditions (8 M urea). Similar to **b**, the C17:0-azide was 'clicked' onto a biotinylated alkyne, and the labelled proteins were pulled-down with streptavidin beads. After washing, immunoprecipitated proteins were eluted off beads in Laemmli buffer containing biotin, and analysed by immunoblotting. The palmitic acid analogue C15:0-azide was used as a positive control since TFR1 is known to also be palmitoylated. C17:0-azide pulls down more TFR1 than equal amounts of C15:0-azide, indicating that TFR1 palmitoylation cannot account for the C17:0 signal. The C17:0-azide-TFR1 interaction is completely blunted upon ZDHHC6 knockdown. See Supplementary Fig. 18 for image of the uncropped full western blot. **d**, TFR1 is required for C18:0 removal to induce mitochondrial fragmentation. HeLa cells were transfected with either control or TFR1 targeting siRNAs before treatment with medium containing delipidated serum plus or minus C18:0. Representative images are shown here and quantification of mitochondrial fragmentation is shown in Fig. 3h. $n = 15$.



Extended Data Figure 8 | C18:0 removal does not affect iron uptake or delivery.

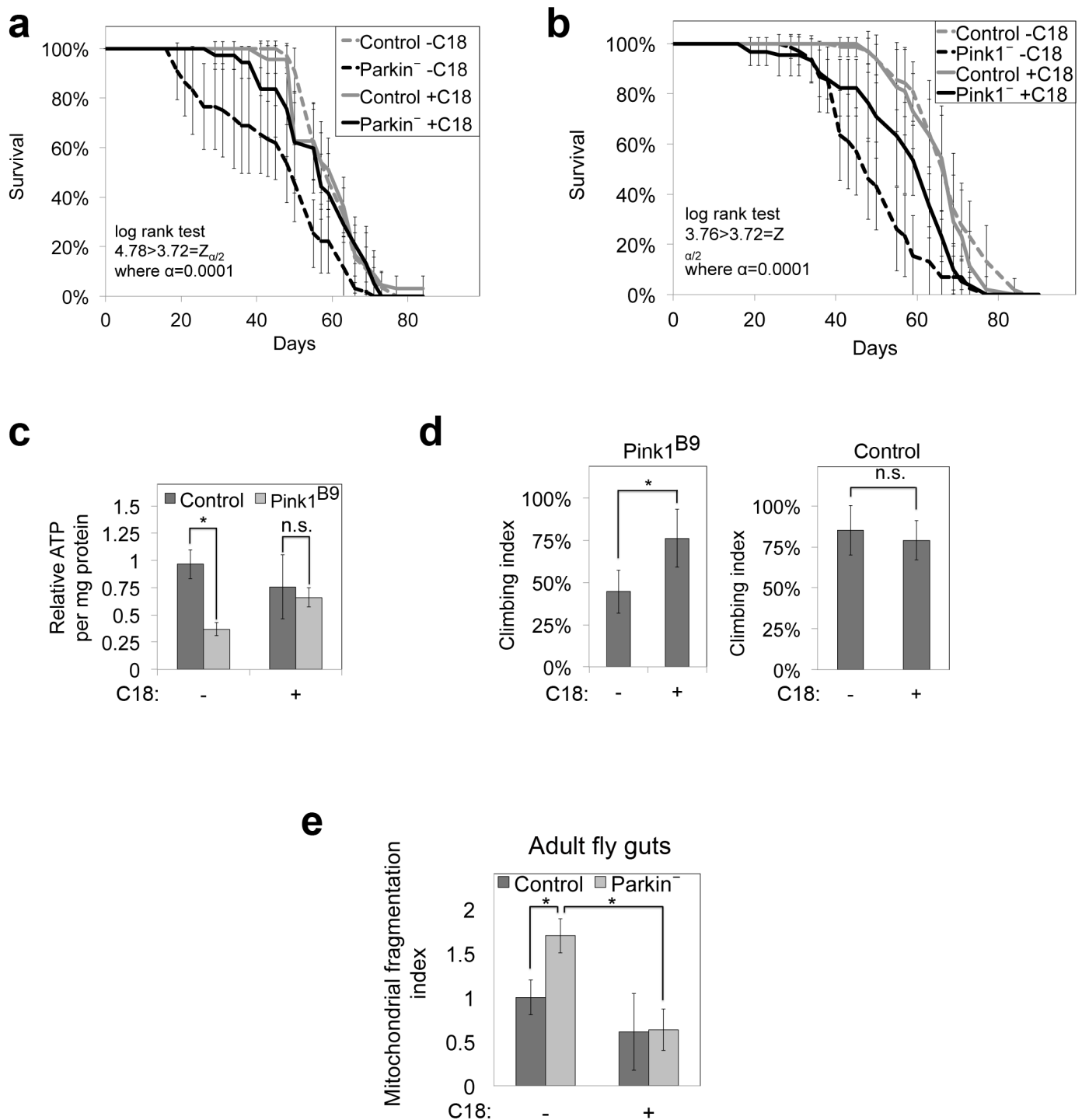
a, HeLa cells cannot grow in the presence of deferoxamine (DFO), an iron chelator (top) whereas they grow in delipidated serum lacking C18:0 at a comparable rate to cells in control medium (bottom). $n = 3$. **b**, Treatment of HeLa cells with medium containing delipidated serum (lacking C18:0) for 24 h does not lead to transcriptional activation of iron deficiency response genes (bottom), which are activated by DFO-mediated iron chelation (24 h) as a positive control (top). $n = 3$. **c**, Treatment of HeLa cells with medium containing delipidated serum for 24 h or 4 days does not lead to a drop in levels of succinate dehydrogenase b (SDHB), which contains an Fe-S cluster. See Supplementary Fig. 19 for image of the uncropped full western blot. **d–f**, Treatment of HeLa cells with medium containing delipidated serum for 24 h or 4 days does not lead to a drop in activities of enzymes containing lipoylated subunits (PDH and OGDH) (**d**, **e**) or Fe-S-cluster-containing

subunits (SDH) (**f**). DFO treatment to chelate iron from the medium, or siRNA-mediated depletion of the enzymes were used as positive controls (**d–f**, bottom). $n = 4$. **g**, Treatment of HeLa cells with medium containing delipidated serum (24 h) does not cause a reduction in transferrin uptake. Cells were treated with $25 \mu\text{g ml}^{-1}$ Alexa-488-coupled transferrin for 30 min. Representative images (left) and quantification of the amount of transferrin per cell (right) ($n = 5$). **h**, Treatment of HeLa cells with medium containing delipidated serum (24 h) does not reduce association of transferrin-containing vesicles with mitochondria. Crude mitochondria were fractionated from cells growing in medium containing or lacking C18:0, and the amount of transferrin that copurifies with mitochondria was analysed and quantified by immunoblotting. See Supplementary Fig. 19 for image of the uncropped full western blot. **a**, **b**, **d–g**, $*P < 0.05$, $***P < 0.001$, $****P < 0.0001$, not significant (NS) $P \geq 0.05$, two-tailed t -test. Error bars show s.d.



Extended Data Figure 9 | JNK signalling is required for mitochondrial fragmentation induced by C18:0 removal. **a**, Treatment of HeLa cells with 1 μ M gambogic acid does not induce apoptosis. 10 μ M gambogic acid was used as a positive control for apoptosis induction, assessed by cleaved caspase-3 levels. 1 μ M gambogic acid neither induces caspase cleavage (shown here) nor causes cells to die (data not shown). Cells were treated with 10 μ M gambogic acid for 1 h, or for all other concentrations for 3 h. See Supplementary Fig. 20 for image of the uncropped full western blot. **b**, Activation of TFR1 by treating cells with 1 μ M gambogic acid leads to mitochondrial fragmentation that is reversed by 1 h C18:0 pre-treatment. Representative images are shown here and quantification of mitochondrial fragmentation is shown in Fig. 3i ($n = 15$). **c**, Treatment of HeLa cells with C18:0 to inhibit TFR1

causes reduced JNK signalling activity, assayed by p-JNK levels on an immunoblot. See Supplementary Fig. 20 for image of the uncropped full western blot. **d**, Removal of C18:0, as well as treatment with gambogic acid, induces shuttling of phosphorylated JNK into the nucleus. Cells were stained with phospho-JNK antibody (left) and relative levels of nuclear to cytosolic phospho-JNK signal was quantified (right) ($n = 37$ cells). *** $P < 0.001$, two-tailed t -test. Error bars show s.d. **e**, JNK signalling is required for TFR1 activation to induce mitochondrial fragmentation. HeLa cells were treated with the JNK inhibitor SP600125 30 min before gambogic acid treatment to activate TFR1. Representative images are shown here and quantification of mitochondrial fragmentation is shown in Fig. 3j ($n = 15$).



Extended Data Figure 10 | Dietary C18:0 improves Parkinson's disease phenotypes of *Pink* and *Parkin* mutant flies. **a, b**, Dietary C18:0 supplementation (10%) significantly increases lifespan of male *Parkin*²⁵ (**a**) and *Pink1*^{B9} (**b**) mutant flies. $n = 8 \times 10$ animals. **c**, Dietary C18:0 supplementation rescues ATP levels of 1-week-old male *Pink1*^{B9} mutant adult flies. $n = 3 \times 3$ animals. **d**, Dietary C18:0 supplementation significantly improves locomotor defects of 2-week-old male *Pink1*^{B9} mutant flies. Locomotion quantified as animals climbing up past a threshold in a given amount of time (technical duplicates, biological quadruplicates, ten animals per assay). **e**, *Parkin* loss of

function in flies leads to mitochondrial fragmentation, which is rescued by dietary supplementation with C18:0. Guts from 14-day-old female control or *park*²⁵ mutant adult flies expressing mitoGFP and grown on food supplemented with or without C18:0 (10%) were dissected and mitochondria were imaged. Quantification of mitochondrial fragmentation is shown (3 animals per condition, 6 optical areas per animal). **b–d**, Control flies are the revertant line *Pink1*^{RV}. Error bars show s.d. Not significant (NS) $P \geq 0.05$, $*P < 0.05$.

GGGGCC repeat expansion in *C9orf72* compromises nucleocytoplasmic transport

Brian D. Freibaum^{1*}, Yubing Lu^{2*}, Rodrigo Lopez-Gonzalez², Nam Chul Kim¹, Sandra Almeida², Kyung-Ha Lee¹, Nisha Badders¹, Marc Valentine¹, Bruce L. Miller³, Philip C. Wong⁴, Leonard Petrucelli⁵, Hong Joo Kim¹, Fen-Biao Gao² & J. Paul Taylor⁶

The GGGGCC (G_4C_2) repeat expansion in a noncoding region of *C9orf72* is the most common cause of sporadic and familial forms of amyotrophic lateral sclerosis and frontotemporal dementia^{1,2}. The basis for pathogenesis is unknown. To elucidate the consequences of G_4C_2 repeat expansion in a tractable genetic system, we generated transgenic fly lines expressing 8, 28 or 58 G_4C_2 -repeat-containing transcripts that do not have a translation start site (AUG) but contain an open-reading frame for green fluorescent protein to detect repeat-associated non-AUG (RAN) translation. We show that these transgenic animals display dosage-dependent, repeat-length-dependent degeneration in neuronal tissues and RAN translation of dipeptide repeat (DPR) proteins, as observed in patients with *C9orf72*-related disease. This model was used in a large-scale, unbiased genetic screen, ultimately leading to the identification of 18 genetic modifiers that encode components of the nuclear pore complex (NPC), as well as the machinery that coordinates the export of nuclear RNA and the import of nuclear proteins. Consistent with these results, we found morphological abnormalities in the architecture of the nuclear envelope in cells expressing expanded G_4C_2 repeats *in vitro* and *in vivo*. Moreover, we identified a substantial defect in RNA export resulting in retention of RNA in the nuclei of *Drosophila* cells expressing expanded G_4C_2 repeats and also in mammalian cells, including aged induced pluripotent stem-cell-derived neurons from patients with *C9orf72*-related disease. These studies show that a primary consequence of G_4C_2 repeat expansion is the compromise of nucleocytoplasmic transport through the nuclear pore, revealing a novel mechanism of neurodegeneration.

To explore pathogenic mechanisms of disease initiated by *C9orf72* repeat expansion in a genetically tractable model organism, we used PhiC31 integrase-mediated insertion of 8, 28 or 58 copies of G_4C_2 repeats ((G_4C_2)₈, (G_4C_2)₂₈ and (G_4C_2)₅₈, respectively) into specific genomic loci of *Drosophila*. These repeats are not expressed at baseline but are transcribed when *GAL4* is introduced *in trans* by genetic cross (Fig. 1a). The green fluorescent protein (GFP) coding sequence without a start codon was placed in frame with a potential poly-(glycine-proline) (poly(GP)) dipeptide repeat (Fig. 1a), although RAN translation of the sense strand may also occur in two other reading frames producing poly(glycine-arginine) (poly(GR)) and poly(glycine-alanine) (poly(GA)).

We expressed these constructs in the *Drosophila* eye using *GMR-GAL4* and observed a length- and dosage-dependent rough eye phenotype (Fig. 1b) similar to that reported recently^{3,4}. Toxicity was also observed when these repeats were expressed in other tissues. Expression of (G_4C_2)₅₈, but not (G_4C_2)₈, in motor neurons using *OK371-GAL4* led to small larvae with significantly impaired locomotor activity (Fig. 1c, d). The neuromuscular junctions (NMJs) were examined with presynaptic and postsynaptic markers (Fig. 1e). Expression of

(G_4C_2)₅₈ resulted in a significant decrease in bouton number and total muscle area when compared with GFP or (G_4C_2)₈ (Fig. 1f). Additionally, active zones within the NMJ were markedly reduced in larvae expressing (G_4C_2)₅₈ compared with controls (Extended Data

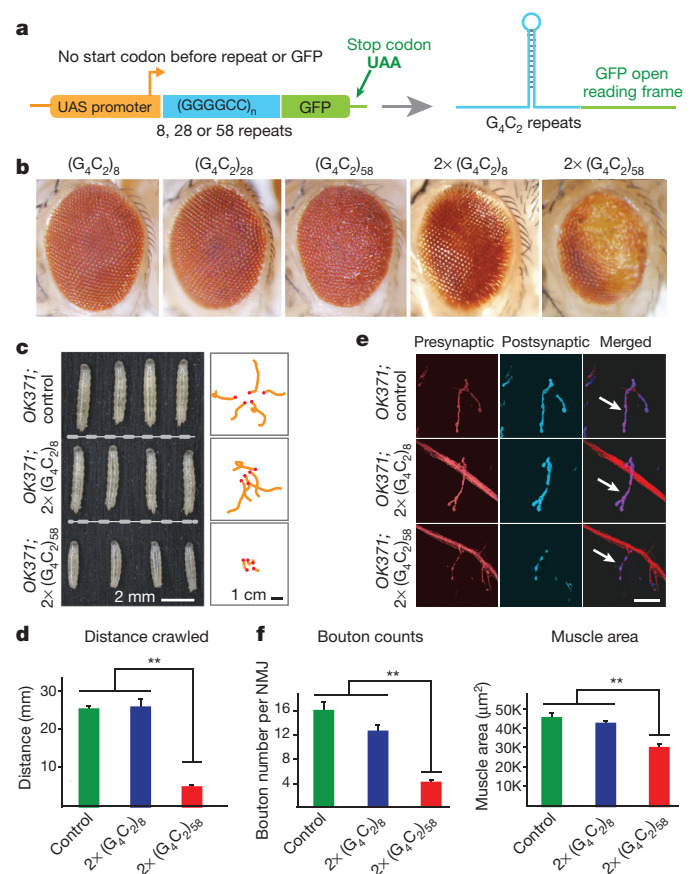


Figure 1 | G_4C_2 repeats induces length- and dosage-dependent degeneration in *Drosophila*. **a**, Constructs expressing 8, 28, or 58 copies of G_4C_2 repeats. **b**, (G_4C_2)₅₈ causes a rough eye phenotype. **c**, Two copies of (G_4C_2)₅₈ expressed in motor neurons result in a decrease in larval size (left) and locomotor activity (right). **d**, Distance travelled by larvae expressing repeats in motor neurons. Values are mean ± s.e.m., n = 5 (control), (G_4C_2)₈ or 6 ($G_4C_2)₅₈ larvae (3 trials per larva), **P < 0.01, by one-way analysis of variance (ANOVA), Tukey's post hoc test. **e**, Expression of two copies of (G_4C_2)₅₈ in motor neurons reduces bouton number. Scale bar, 25 μm. **f**, Bouton number and muscle size in larvae expressing G_4C_2 repeats. Values are mean ± s.e.m., n = 6 larvae (3 trials per larva), **P < 0.01, by one-way ANOVA, Tukey's post hoc test.$

¹Department of Cell and Molecular Biology, St Jude Children's Research Hospital, Memphis, Tennessee 38105, USA. ²Department of Neurology, University of Massachusetts Medical School, Worcester, Massachusetts 01605, USA. ³Memory and Aging Center, Department of Neurology, University of California San Francisco, San Francisco, California 94158, USA. ⁴Department of Pathology, The Johns Hopkins University School of Medicine, Baltimore, Maryland 21205, USA. ⁵Department of Neuroscience, Mayo Clinic Florida, Jacksonville, Florida 32224, USA. ⁶Howard Hughes Medical Institute, Department of Cell and Molecular Biology, St Jude Children's Research Hospital, Memphis, Tennessee 38105, USA.

*These authors contributed equally to this work.

In this fly model, RAN translation occurs in a G_4C_2 -repeat-length-dependent manner. Poly(GP)–GFP dipeptide repeats were detected in flies expressing $(G_4C_2)_{58}$ and to a lesser extent $(G_4C_2)_{28}$, but not in

flies expressing equivalent levels of (G₄C₂)₈ (Extended Data Fig. 2a, b). Indeed, nuclear and cytoplasmic inclusions of poly(GP)–GFP were present in tissues such as brain, muscle and salivary gland expressing (G₄C₂)₅₈ but not (G₄C₂)₈ (Extended Data Fig. 2c–g). To determine whether DPRs are sufficient to drive a degenerative phenotype, we used codons alternative to G₄C₂ to generate new transgenic flies directly expressing poly(GA), poly(GR), or poly(GP), with an AUG start codon and amino-terminal GFP (Supplementary Table 1 and Extended Data Fig. 3a). Expression of neither GFP–(GP)₄₇ nor

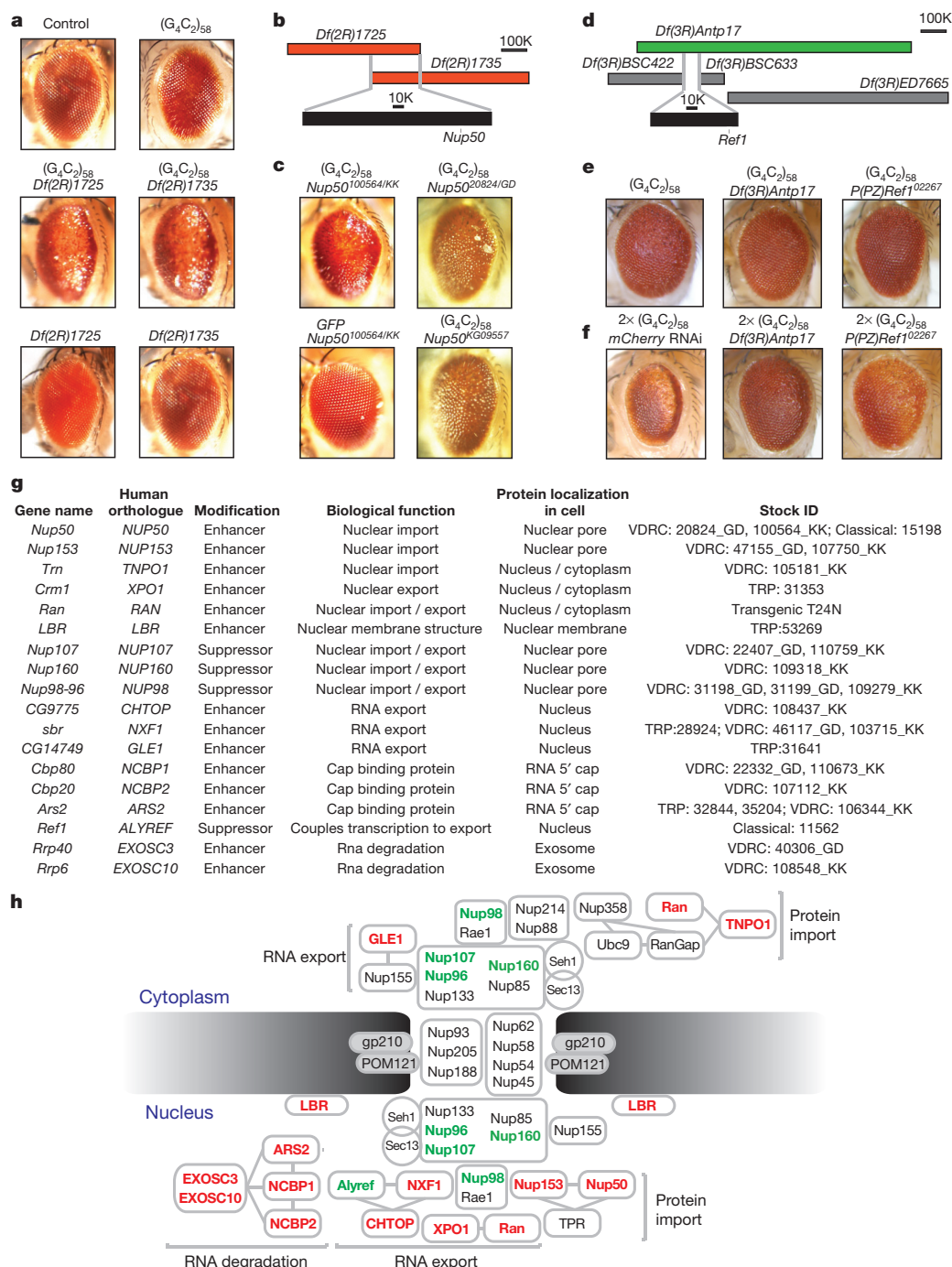


Figure 2 | Genetic screen identifies multiple modifiers of (G₄C₂)₅₈ toxicity in the nucleocytoplasmic transport pathway. **a**, (G₄C₂)₅₈ expression driven by *GMR-GAL4* causes rough eye phenotypes (top right) which are enhanced by either *Df(2R)1725/+* or *Df(2R)1735/+* (middle). **b**, Overlapping genomic region of deficiency lines *Df(2R)1725* or *Df(2R)1735*. **c**, RNAi knockdown (*Nup50*^{20824/GD} and *Nup50*^{100564/KK}) and genetic allele (*Nup50*^{KG09557}) identify

Nup50 as an enhancer. **d**, The deficiency *Df(3R)Antp17* (green) but not others (grey) suppressed the $(G_4C_2)_{58}$ phenotype. **e, f**, Identification of *Ref1* as a suppressor. **g**, Table summarizing modifiers and their functions. **h**, Suppressors (green) and enhancers (red) of $(G_4C_2)_{58}$ toxicity in the nucleocytoplasmic trafficking pathway.

GFP-(GA)₅₀ elicited a degenerative eye phenotype (Extended Data Fig. 3a, b). Thus, while poly(GP)-GFP serves as a useful marker of cells expressing (G₄C₂)₅₈ RNA and producing DPR proteins, it is unlikely to contribute substantially to the degenerative phenotype. By contrast, expression of GFP-(GR)₅₀ was highly toxic, resulting in greater than 95% lethality with a few escapers having severely degenerated eyes (Extended Data Fig. 3a). These results are consistent with recent reports indicating that poly(GR) is toxic in cultured cells and *Drosophila*^{3,5,6}. In fact, immunoblotting with antibodies against DPRs showed that flies expressing (G₄C₂)₅₈ RNA produce poly(GR) in addition to poly(GP) DPRs, although neither poly(GA) nor poly(PR) was detected (Extended Data Fig. 3c, d). Thus, (G₄C₂)₅₈ toxicity may be mediated through toxic RNA or poly(GR) produced in our fly model, or a combination of these mechanisms.

To gain unbiased insight into the pathogenic mechanisms underlying (G₄C₂)₅₈ toxicity, we performed a large-scale genetic screen to identify genetic loci whose partial loss of function may strongly modify the (G₄C₂)₅₈-induced rough eye phenotype. To this end, flies expressing (G₄C₂)₅₈ in the eye were crossed with 372 chromosomal deficiency lines spanning the entirety of the second and third chromosomes, representing ~80% of the fly genome. In total, 5.1% of the deficiencies were found to suppress the rough eye phenotype, whereas 8.8% of these enhanced the rough eye phenotype (Supplementary Table 2).

Two deficiency lines on the second chromosome, *Df(2R)1725* and *Df(2R)1735*, each in heterozygotes, had no eye phenotype by themselves but led to a marked enhancement of the (G₄C₂)₅₈ rough eye phenotype (Fig. 2a). The genomic regions covered by these deficiencies partially overlap (Fig. 2b), suggesting that one or more genes in this overlapping region are responsible for the observed enhancement. After systematic evaluation of this candidate interval with classical loss of function (LOF) alleles and RNA interference (RNAi) lines, we identified the gene *Nup50*, the partial LOF of which enhanced the (G₄C₂)₅₈ phenotype (Fig. 2c). The effect of *Nup50* was specific to (G₄C₂)₅₈, since there was no rough eye phenotype present when *Nup50* was knocked down in eyes expressing GFP (Fig. 2c). *Nup50* is a component of the nuclear pore and also has a critical role in promoting protein nuclear import through interaction with Importin β and Ran GTPase⁷. Consistent with the impact of *Nup50* LOF on (G₄C₂)₅₈-mediated degeneration, a dominant-negative form of Ran (Ran^{T24N})⁸ strongly enhanced the (G₄C₂)₅₈ rough eye phenotype (Extended Data Fig. 4a). Moreover, the nuclear import factors *Nup153* and *Transportin*, which work together with *Nup50* and Ran⁷, were also identified as enhancers of the (G₄C₂)₅₈ rough eye phenotype (Extended Data Fig. 4a). These genetic analyses suggest that protein import is compromised by (G₄C₂)₅₈ expression.

The strongest suppressor identified in the genetic screen was *Ref1*, revealed by interrogation of the suppressor deficiency *Df(3R)Antp17* (Fig. 2d, e). This deficiency also suppressed the strong rough eye phenotype produced by two copies of (G₄C₂)₅₈, as did heterozygosity for the null allele *Ref1*⁰²²⁶⁷ (Fig. 2f). *Ref1*, and its human orthologue *ALYREF*, are RNA-binding proteins that associate with the 5' end of messenger RNAs to prevent degradation by the nuclear exosome and, together with the TREX component CHTOP, facilitate delivery of fully processed mRNAs to the nuclear pore receptor NXF1, which mediates their export through the nuclear pore^{9–12}. Consistent with the impact of *Ref1* LOF on (G₄C₂)₅₈-mediated degeneration, partial LOF in *Drosophila* orthologues of *NXF1* or *CHTOP* enhanced the rough eye phenotype (Fig. 2g, h and Extended Data Fig. 5). The fact that the (G₄C₂)₅₈ phenotype is exacerbated by LOF in *NXF1* or *CHTOP* orthologues, but partially rescued by LOF of *Ref1*, perhaps highlights the dual role played by *Ref1* in partitioning target RNAs between the nuclear pore and the nuclear exosome⁹. In support of this interpretation, we identified LOF in *Drosophila* orthologues of human *EXOSC3* and *EXOSC10*, which encode components of the nuclear exosome, as strong enhancers of (G₄C₂)₅₈-related toxicity (Fig. 2g, h and Extended

Data Fig. 5). Notably, LOF mutations in *EXOSC3* in humans cause a congenital form of motor neuron disease¹³.

The *Drosophila* genes encoding human orthologues of cap-binding proteins NCBP1, NCBP2 and ARS2, which mediate recruitment of the TREX complex to the 5' end of RNA to initiate RNA export¹⁴, were also identified as enhancers of (G₄C₂)₅₈-mediated degeneration (Fig. 2g, h and Extended Data Fig. 5). Moreover, we found that LOF of either *Nup107* or its binding partner *Nup160*, both of which are nuclear pore components responsible for exporting some RNAs^{15–17}, suppressed the (G₄C₂)₅₈ rough eye phenotype (Extended Data Fig. 5). One particularly notable enhancer was the *Drosophila* orthologue of human *GLE1*, a critical mediator of RNA export at the nuclear pore¹⁸ (Fig. 2g, h and Extended Data Fig. 5). Interestingly, complete LOF of human *GLE1* causes congenital motor neuron disease¹⁹, whereas partial LOF of human *GLE1* is associated with adult-onset amyotrophic lateral sclerosis (ALS)²⁰.

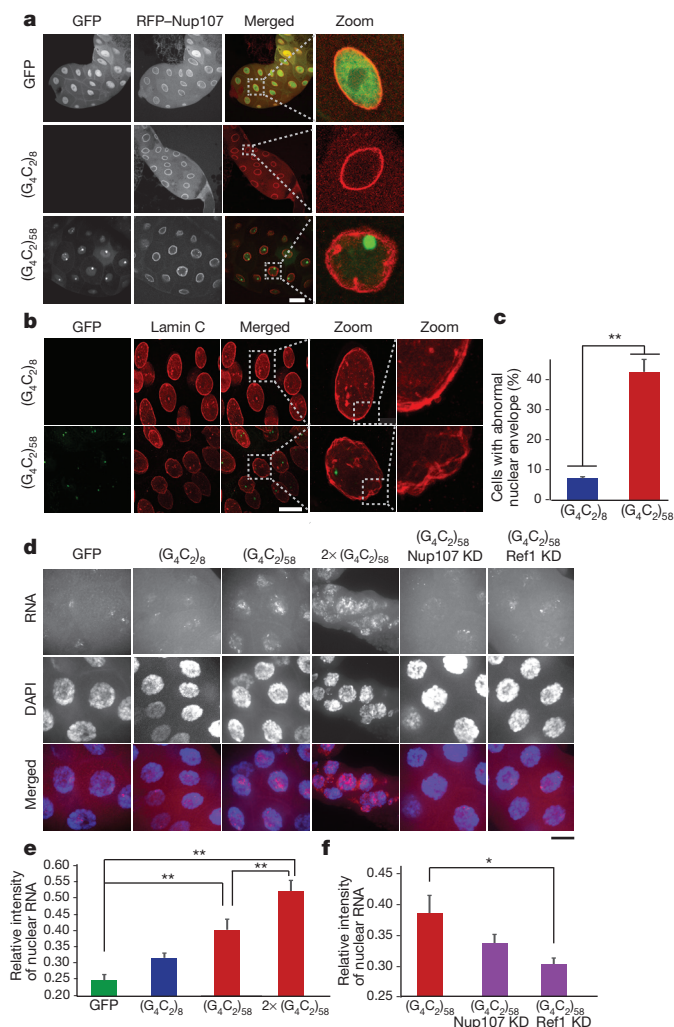


Figure 3 | *Drosophila* salivary gland cells expressing (G₄C₂)₅₈ exhibit nuclear envelope abnormalities and accumulation of nuclear RNA. **a**, The effect of (G₄C₂)₅₈ expression on Nup107 localization. Scale bar, 50 μ m. **b**, (G₄C₂)₅₈ expression causes abnormal nuclear envelope morphology. Scale bar, 50 μ m. **c**, Nuclear envelope phenotype in cells expressing either (G₄C₂)₅₈ ($n = 251$ cells) or (G₄C₂)₅₈ Ref1 KD ($n = 127$ cells). Values are mean \pm s.e.m. from 3 biological replicates. **d**, Accumulation of total nuclear RNA relative to cytoplasmic RNA (red). Knockdown (KD) of *Ref1* led to a partial rescue of nuclear RNA accumulation. Scale bar, 25 μ m. **e**, **f**, Relative intensity of total nuclear RNA versus cytoplasmic RNA in cells from random fields of 5 larvae per genotype. Values are mean \pm s.e.m., $n = 30$ cells (60 cells for (G₄C₂)₅₈/Nup107 KD and (G₄C₂)₅₈/Ref1 KD). ***P** < 0.01, ****P** < 0.001 by one-way ANOVA, Tukey's post hoc test.

The identification of TREX and nuclear pore proteins as genetic modifiers strongly implies that not only is nuclear import compromised by $(G_4C_2)_{58}$ expression but also that nuclear export of RNAs and proteins is compromised. Consistent with this latter notion we found that LOF of *Crml* (also called *emb*), a gene encoding a major receptor for the export of RNAs and proteins, resulted in strong enhancement of the $(G_4C_2)_{58}$ phenotype (Extended Data Fig. 4b). Moreover, expression of $(G_4C_2)_{58}$ in motor neurons caused exquisite sensitivity to leptomycin B, an inhibitor of nuclear export (Extended Data Fig. 4c). In total, the genetic screen and follow-up analyses identified 18 modifier genes within the pathway of nucleocytoplasmic transport, and RNA export in particular, highlighting this system as an important target of $(G_4C_2)_{58}$ -related toxicity (Fig. 2g, h and Supplementary Table 3). In particular, the identification of four strong suppressors of $(G_4C_2)_{58}$ indicates that compromised nucleocytoplasmic transport is an important causal pathway responsible for degeneration.

In follow up to our genetic studies, RFP-Nup107 was expressed in salivary glands using *Fkh-GAL4*. In cells expressing $(G_4C_2)_8$, Nup107 labels a distinct nuclear boundary that is morphologically indistinguishable from cells expressing GFP (Fig. 3a). In contrast, in cells expressing $(G_4C_2)_{58}$, the nuclear envelope exhibited a wrinkled appearance and in many nuclei Nup107 was found to form inclusions near the nuclear envelope (Fig. 3a). To explore nuclear architecture further, the nuclear envelope was visualized by immunostaining of endogenous Lamin C. A total of 43.7% of cells expressing $(G_4C_2)_{58}$ showed an abnormal, 'frayed' nuclear envelope phenotype, whereas this phenotype was observed in only 7.1% of cells expressing $(G_4C_2)_8$ (Fig. 3b, c and Extended Data Fig. 6a). Thus, $(G_4C_2)_{58}$ expression causes defects in the architecture of the nuclear envelope and mislocalization of nucleoporins. Moreover, cells expressing one copy of $(G_4C_2)_{58}$ showed a significant increase in the ratio of nuclear to cytoplasmic RNA in comparison to cells expressing GFP or $(G_4C_2)_8$ (Fig. 3d, e). Expression of two copies of $(G_4C_2)_{58}$ resulted in a more pronounced retention of nuclear RNA (Fig. 3d, e). Importantly, depletion

of Ref1 by RNAi partially suppressed the nuclear RNA retention phenotype in cells expressing $(G_4C_2)_{58}$, reducing nuclear RNA density to control levels (Fig. 3d–f). Thus, G_4C_2 length-dependent degeneration in *Drosophila* is accompanied by nuclear retention of RNA, consistent with the results of the unbiased genetic screen.

Similar retention of nuclear RNA was observed in HeLa and HEK293T cells upon transient expression of $(G_4C_2)_{58}$ but not $(G_4C_2)_8$ (Fig. 4a–e and Extended Data Fig. 6b). Specifically, this export defect was illustrated using an alternative approach to monitor the fate of newly synthesized RNA by metabolic labelling. In HeLa cells expressing $(G_4C_2)_{58}$ we observed a significant decrease in the export of nascent RNA in comparison to control cells expressing GFP or $(G_4C_2)_8$, although the total levels of metabolically labelled transcripts were comparable (Fig. 4a–c and Extended Data Fig. 6c). Moreover, poly(A)⁺ mRNA as detected by an oligo-dT probe accounts for at least some portion of the retained RNA in mammalian cells expressing $(G_4C_2)_{58}$ (Fig. 4d, e).

Advances in induced pluripotent stem cell (iPSC) technology permit the generation of neurons from patients including those with frontotemporal dementia (FTD) and ALS²¹. To examine further the impact of expanded G_4C_2 toxicity on nucleocytoplasmic transport in a more relevant human cell type from *C9orf72* patients, we examined the subcellular distribution of total RNA in 2-month-old cortical neurons, which are affected in FTD, a frequent clinical feature of *C9orf72*-related disease²². These neurons were differentiated from previously characterized^{23–25} as well as newly generated and characterized iPSC lines derived from 5 *C9orf72* patients and 3 controls (Extended Data Figs 7 and 8 and Supplementary Tables 4 and 5). Consistent with our observations in fly, HEK293T and HeLa cells, this analysis confirmed a significant increase in the nuclear to cytoplasmic ratio of RNA in *C9orf72* patient cells in comparison to controls (Fig. 4f, g). On average, *C9orf72* neurons showed a 35% increase in the nuclear to cytoplasmic ratio of RNA density (Fig. 4g). Notably, this difference in the ratio of nuclear to cytoplasmic RNA was not observed in fibroblasts obtained from *C9orf72* patients when compared to controls, consistent with low

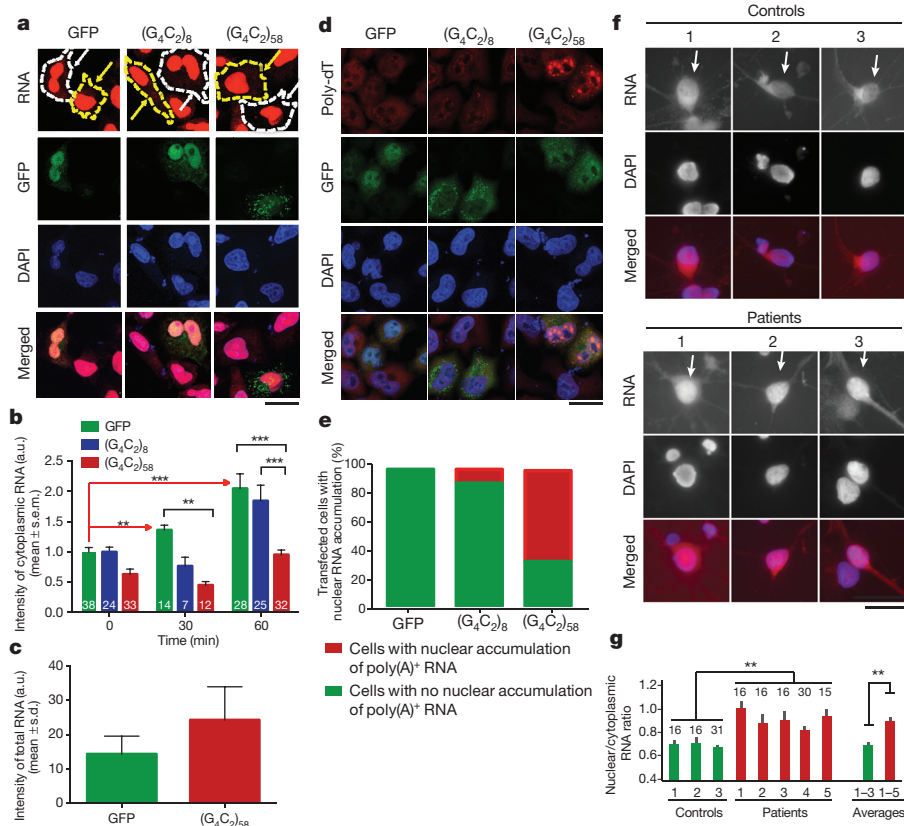


Figure 4 | Accumulation of nuclear RNA in human cells expressing G_4C_2 repeat expansion.

a, Reduced accumulation of newly synthesized RNA in the cytoplasm of cells transfected with $(G_4C_2)_{58}$ (white arrows) (imaged at 60 min). Yellow arrows indicate untransfected cells. Scale bar, 20 μ m. **b**, Cytoplasmic RNA intensity after metabolic labelling of newly synthesized RNA. The number of cells quantified per genotype (5 biological replicates) is listed on the graph. **c**, Total RNA intensity after labelling of newly synthesized RNA at time point 0 min. **d**, Poly(A)⁺ RNA was measured by FISH in transiently transfected HeLa cells. Scale bar, 20 μ m. **e**, Cells containing poly(A)⁺ RNA puncta, $n = 30$ cells per condition (2 biological replicates). **f**, Total cellular RNA in 2-month-old iPSC-derived cortical neurons. Scale bar, 25 μ m. **g**, Cytoplasmic RNA ratio in 5 patient cortical neurons versus 3 controls, cells per subject is labelled on graph. Values are mean \pm s.e.m., $**P < 0.01$ by Student's *t*-test.

levels of *C9orf72* expression in fibroblasts²⁵ (Extended Data Fig. 9a–c). Thus, neurons derived from patients with *C9orf72*-related disease mirror the nucleocytoplasmic transport defects identified by unbiased approaches in a genetic model of disease.

The NPC is the largest macromolecular complex in eukaryotic cells, consisting of multiple copies of more than 30 different proteins²⁶. Some structural components of the NPC have an exceedingly long half-life measured in years²⁷. Indeed, nuclear pores are a particularly intriguing target for age-related diseases affecting post-mitotic cells such as neurons, putting them at risk of accumulating damage over extended periods of time²⁷. Nucleocytoplasmic transport through the nuclear pore is vital to cell viability, and primary defects in this function cause a host of human diseases of varying phenotypes ranging from cancer to neurological disease^{26,28,29}.

Important questions regarding the pathogenesis of *C9orf72*-related disease remain; in particular the relative contribution of expanded G₄C₂ RNA versus DPRs or other mechanisms, which could cause the nuclear pore defect. Also to be determined is whether defects in nucleocytoplasmic transport and other reported defects such as sequestration of specific RNA binding proteins, nucleolar stress and ER stress are parallel processes or directly connected with each other. The rapid pace of discovery in this area of disease biology suggests that answers to these questions will come soon.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 22 December 2014; accepted 24 July 2015.

Published online 26 August 2015.

- DeJesus-Hernandez, M. *et al.* Expanded GGGGCC hexanucleotide repeat in noncoding region of C9ORF72 causes chromosome 9p-linked FTD and ALS. *Neuron* **72**, 245–256 (2011).
- Renton, A. E. *et al.* A hexanucleotide repeat expansion in C9ORF72 is the cause of chromosome 9p21-linked ALS-FTD. *Neuron* **72**, 257–268 (2011).
- Mizielinska, S. *et al.* C9orf72 repeat expansions cause neurodegeneration in *Drosophila* through arginine-rich proteins. *Science* **345**, 1192–1194 (2014).
- Xu, Z. *et al.* Expanded GGGGCC repeat RNA associated with amyotrophic lateral sclerosis and frontotemporal dementia causes neurodegeneration. *Proc. Natl Acad. Sci. USA* **110**, 7778–7783 (2013).
- Kwon, I. *et al.* Poly-dipeptides encoded by the C9orf72 repeats bind nucleoli, impede RNA biogenesis, and kill cells. *Science* **345**, 1139–1145 (2014).
- Wen, X. *et al.* Antisense proline-arginine RAN dipeptides linked to C9ORF72-ALS/FTD form toxic nuclear aggregates that initiate *in vitro* and *in vivo* neuronal death. *Neuron* **84**, 1213–1225 (2014).
- Makise, M. *et al.* The Nup153-Nup50 protein interface and its role in nuclear import. *J. Biol. Chem.* **287**, 38515–38522 (2012).
- Cesario, J. & McKim, K. S. RanGTP is required for meiotic spindle organization and the initiation of embryonic development in *Drosophila*. *J. Cell Sci.* **124**, 3797–3810 (2011).
- Reed, R. Coupling transcription, splicing and mRNA export. *Curr. Opin. Cell Biol.* **15**, 326–331 (2003).
- Viphakone, N. *et al.* TREX exposes the RNA-binding domain of Nxf1 to enable mRNA export. *Nat. Commun.* **3**, 1006 (2012).
- Zhou, Z. *et al.* The protein Aly links pre-messenger-RNA splicing to nuclear export in metazoans. *Nature* **407**, 401–405 (2000).
- Chang, C. T. *et al.* Chtop is a component of the dynamic TREX mRNA export complex. *EMBO J.* **32**, 473–486 (2013).
- Wan, J. *et al.* Mutations in the RNA exosome component gene EXOSC3 cause pontocerebellar hypoplasia and spinal motor neuron degeneration. *Nature Genet.* **44**, 704–708 (2012).
- Cheng, H. *et al.* Human mRNA export machinery recruited to the 5' end of mRNA. *Cell* **127**, 1389–1400 (2006).
- Boehmer, T., Enninga, J., Dales, S., Blobel, G. & Zhong, H. Depletion of a single nucleoporin, Nup107, prevents the assembly of a subset of nucleoporins into the nuclear pore complex. *Proc. Natl Acad. Sci. USA* **100**, 981–985 (2003).
- Vasu, S. *et al.* Novel vertebrate nucleoporins Nup133 and Nup160 play a role in mRNA export. *J. Cell Biol.* **155**, 339–354 (2001).
- Walther, T. C. *et al.* The conserved Nup107–160 complex is critical for nuclear pore complex assembly. *Cell* **113**, 195–206 (2003).
- Murphy, R. & Wentz, S. R. An RNA-export mediator with an essential nuclear export signal. *Nature* **383**, 357–360 (1996).
- Nousiainen, H. O. *et al.* Mutations in mRNA export mediator GLE1 result in a fetal motoneuron disease. *Nature Genet.* **40**, 155–157 (2008).
- Kaneb, H. M. *et al.* Deleterious mutations in the essential mRNA metabolism factor, hGle1, in amyotrophic lateral sclerosis. *Hum. Mol. Genet.* **24**, 1363–1373 (2015).
- Han, S. S., Williams, L. A. & Eggan, K. C. Constructing and deconstructing stem cell models of neurological disease. *Neuron* **70**, 626–644 (2011).
- van Blitterswijk, M., DeJesus-Hernandez, M. & Rademakers, R. How do C9ORF72 repeat expansions cause amyotrophic lateral sclerosis and frontotemporal dementia: can we learn from other noncoding repeat expansion disorders? *Curr. Opin. Neurol.* **25**, 689–700 (2012).
- Almeida, S. *et al.* Modeling key pathological features of frontotemporal dementia with C9ORF72 repeat expansion in iPSC-derived human neurons. *Acta Neuropathol.* **126**, 385–399 (2013).
- Zhang, Z. *et al.* Downregulation of microRNA-9 in iPSC-derived neurons of FTD/ALS patients with TDP-43 mutations. *PLoS ONE* **8**, e76055 (2013).
- Sareen, D. *et al.* Targeting RNA foci in iPSC-derived motor neurons from ALS patients with a C9ORF72 repeat expansion. *Sci. Transl. Med.* **5**, 208ra149 (2013).
- Cronshaw, J. M. & Matunis, M. J. The nuclear pore complex: disease associations and functional correlations. *Trends Endocrinol. Metab.* **15**, 34–39 (2004).
- Savas, J. N., Toyama, B. H., Xu, T., Yates, J. R. III & Hetzer, M. W. Extremely long-lived nuclear pore proteins in the rat brain. *Science* **335**, 942 (2012).
- Mor, A., White, M. A. & Fontoura, B. M. Nuclear trafficking in health and disease. *Curr. Opin. Cell Biol.* **28**, 28–35 (2014).
- Bonnet, A. & Palancade, B. Regulation of mRNA trafficking by nuclear pore complexes. *Genes* **5**, 767–791 (2014).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank the Bloomington *Drosophila* Stock Center, the VDRC Stock Center, K. McKim and E. Baehrecke for fly lines, V. Budnik for Lamin C antibody, as well as the Cell and Tissue Imaging Center at St Jude Children's Research Hospital and the University of Massachusetts Medical School Confocal Core for assistance. This work was supported by grants from Target ALS, The Packard Center for ALS Research at the Johns Hopkins University, and the ALS Association to F.-B.G., and J.P.T., and ALS Therapy Alliance, NIH (N079725) to F.-B.G., NIH (NS079725 and AG019724) to B.L.M., and the American-Lebanese-Syrian Associated Charities to J.P.T.

Author Contributions F.-B.G. and J.P.T. conceived and supervised the project. B.D.F., Y.L., H.J.K., F.-B.G. and J.P.T. wrote the manuscript. B.D.F. and Y.L. performed the genetic screen and validation. B.D.F., Y.L., N.C.K., N.B. and K.-H.L. characterized *Drosophila* phenotypes and performed the assays characterizing RNA export in human cells. S.A. established human fibroblast cell lines, R.L.-G. generated some iPSC lines; S.A. and R.L.-G. performed cortical neuron differentiation. M.V. and B.D.F. conducted FISH experiments. L.P., B.L.M. and P.C.W. provided key reagents.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.P.T. (jpaul.taylor@stjude.org) or F.-B.G. (fen-biao.gao@umassmed.edu).

METHODS

The experiments were not randomized, and the investigators were not blinded to allocation during experiments and outcome assessment. No statistical methods were used to predetermine sample size.

Generation of *Drosophila* lines. To generate transgenic *Drosophila* expressing (G_4C_2)₈, (G_4C_2)₂₈ and (G_4C_2)₅₈, G_4C_2 repeats of the respective length were cloned downstream of the UAS promoter and upstream of the EGFP sequence with the start codon removed in the plasmid pUAST-ATTB. Transgenic *Drosophila* lines were generated by BestGene Inc. such that the transgene was inserted using the PhiC31 integrase into either the attP2 site on chromosome 3 (loci 68A4) or the attP40 site on chromosome 2 (loci 25C6). GFP-(GP)₄₇, GFP-(GA)₅₀ and GFP-(GR)₅₀ plasmids³⁰ were subcloned into the pUAST-ATTB plasmid and inserted into the attP2 site by BestGene Inc. Flies were raised at 25 °C on a standard diet. A complete listing of *Drosophila* stocks that modify the (G_4C_2)₅₈ phenotype is given in Supplementary Table 3. For genetic interaction studies, the combined stock (*GMR-GAL4/Cyo*; *UAS-(G₄C₂)₅₈/TM6,Tb*) was crossed with deficiency stocks or individual mutants, and phenotypic analysis was performed on flies aged 24–48 h in both males and females. Knock down efficiency of RNAi lines was measured by quantitative RT-PCR (Extended Data Fig. 10 and Supplementary Table 6). For nuclear envelope morphology studies, the recombined stocks (*Fkh-GAL4, UAS-GFP/TM6,Tb*; *Fkh-GAL4, UAS-(G₄C₂)₈/TM6,Tb*; and *Fkh-GAL4, UAS-(G₄C₂)₅₈/TM6,Tb*) were crossed with *UAS-mRFP-Nup107* transgenic flies or immunostained for endogenous Lamin C.

***Drosophila* stocks.** Deficiency stocks were obtained from the Bloomington *Drosophila* Stock Center deficiency kit. RNAi lines were obtained from either the Bloomington *Drosophila* Stock Center or the Vienna *Drosophila* RNAi Center. *UAS-RAN^{T24N}* and *Fkh-Gal4* flies were provided by K. S. McKim and E. Baehrecke, respectively.

***Drosophila* eye, muscle and neuronal phenotype analysis.** Phenotypic analysis of G_4C_2 repeat expression in the *Drosophila* eye, muscle and salivary gland was assessed by crossing (G_4C_2)₈, (G_4C_2)₂₈ and ($G_4C_2)₅₈ lines to *GMR-GAL4*, *MHC-GAL4* and *Fkh-GAL4*, respectively. Neuronal expression was achieved by crossing (G_4C_2)₈, ($G_4C_2)₂₈ and ($G_4C_2)₅₈ lines to *elav-GAL4* (pan neuronal driver) or *OK371-GAL4* (motor neuron driver). Eye phenotypes were imaged by light microscopy and muscle phenotypes were visually assessed by wing posture ($n = 30$ controls, 21 (G_4C_2)₈, 50 ($G_4C_2)₂₈ and 24 ($G_4C_2)₅₈). Eye phenotypes are representative images resulting from *Drosophila* crosses. Crosses were performed twice to validate the specific phenotype. Neuromuscular synaptic bouton number, active zone density and crawling ability were measured as previously described^{31,32} with slight modification. To count type 1b synaptic bouton number, each genotype was double stained presynaptically with anti-HRP-Cy3 (1:200, Jackson ImmunoResearch) and postsynaptically with anti-Disc large 1 (1:50, DSHB). Synaptic boutons of muscle 4 in abdominal segments 2, 3 and 4 (A2–A4) were imaged with a Marianas spinning disc microscope and maximum projection images were used to count synaptic boutons. Active zone area and presynaptic area stained by anti-Bruchpilot (NC82, 1:100, DSHB) and anti-HRP-Cy3 were measured with ImageJ and the ratio of active zone area/presynaptic area was calculated. Muscle 4 in abdominal segments 2, 3 and 4 (A2–A4) was analysed. To examine crawling ability, 4–7 wandering third instar larvae for each group were collected, washed and placed onto a 3% agarose gel in a 10 cm dish. After 5 min acclimation, larval crawling behaviour was recorded by a digital camera for 30 s (15 frame per second (fps)). Each larva was tested three times (3 technical replicates). Moving distances of each larva were manually measured with ImageJ.$$$$$

Immunofluorescence analysis of salivary gland cells. Dissected salivary glands from wandering third instar larvae were fixed with 4% paraformaldehyde (PFA) at room temperature for 15 min, washed three times with phosphate-buffered saline (PBT, 0.1% Triton X-100 in PBS), and blocked for 1 h at room temperature with 10% normal goat serum (Sigma) diluted in PBT. Tissues were then incubated overnight at 4 °C with mouse anti-Lamin C (1:30, LC28.26; Developmental Studies Hybridoma Bank (DSHB)). After three washes in PBT (10 min each), tissues were incubated for 1 h at room temperature with secondary antibodies (goat anti-mouse Alexa Fluor 568, 1:200; Invitrogen) diluted in the blocking solution. Tissues were washed three times in PBS and mounted with VECTASHIELD (Vector Laboratories). Fluorescence signals were examined with an Olympus IX70 Microscope or Leica TCS SP5 II laser scanning confocal microscope.

Visualization of RAN products in muscle cells. To visualize poly(GP)–GFP in the thorax expressing (G_4C_2)₅₈, flies were cleared overnight using a modified ScaleA2 (2 M urea, 10% glycerol, 0.1% Triton-X). RAN peptides were visualized using light sheet microscopy. To visualize RAN products at high magnification, thoraxes were dissected and stained with the same immunofluorescence protocol described above for salivary glands but using anti-Lamin (1:100, ADL 67.10;

Developmental Studies Hybridoma Bank (DSHB)) to visualize the nucleus and phalloidin (1:40, Life Technologies) to visualize the muscle structure.

Immunoblots. Adult flies were frozen with dry ice and vortexed to remove the head or thorax. Samples from each genotype were homogenized in RIPA buffer with proteinase inhibitor cocktail added. Sample was mixed with 4× sample buffer (1 M Tris-HCl (pH 6.8), 8% SDS, 40% glycerol, 0.1% bromophenol blue) and boiled for 5 min, separated on an SDS gel and transferred to a membrane. The membrane was blocked, probed with primary antibody, and incubated with secondary antibody. The signal was visualized with either chemiluminescent substrate (SuperSignal West Pico; Pierce) or by using an Odyssey Fc (Li-Cor). RAN products were visualized by dot blot using the previously described protocol³. Primary western blot antibodies were anti-GFP (AB3080, Millipore or SC-9996 Santa Cruz Biotechnology), anti-poly(GP), poly(GA) or poly(GR) antibodies^{33,34}, anti-β-actin antibody (4967; Cell Signaling Technology) or anti-actin antibody (sc-1616, Santa Cruz Biotechnology).

RNA *in situ* hybridization. Fixation of dissected salivary glands, cortical neurons, HeLa, or 293T cells (obtained from ATCC) was performed using 1% paraformaldehyde plus 0.05% NP40 for an additional 5 min. Samples were then transferred to 70% ethanol and stored at –20 °C until needed. The probes for global RNA in either *Drosophila* or human cells were prepared from genomic DNA derived from w^{1118} flies or 293T cells. Both probes were labelled by nick translation using either alexafluor 594 dUTP. To hybridize the probes, 70 ng of the labelled probes (the human probe was combined with 2 μg of human cot1 DNA) were suspended in 10 μl of hybridization buffer consisting of 50% formamide, 2× SSC, and 10% dextran sulfate. Fixed cells were dehydrated in 70%, 80% and 100% ethanol for 2 min each before hybridization. Probes suspended in hybridization buffer were denatured at 70 °C for 5 min and then applied to the dehydrated cells and hybridized at 37 °C overnight. Samples were washed after hybridization in 50% formamide, 2× SSC at 37 °C for 5 min, then briefly rinsed in room temperature PBS and mounted on slides with DAPI. To detect poly(A) mRNA, HeLa cells were fixed with 4% paraformaldehyde followed by ice cold methanol and 70% ethanol for 10 min each. The cells were treated with 1 ng per 1 μl of 5′-labelled Cy3-oligo-dT(20) at 37 °C for 1 h. Samples were washed in room temperature PBS and mounted on slides with DAPI. Images were captured using either wide-field fluorescence microscopy or a Marianas spinning disc microscope. An extended depth of focus function was used to combine all components of individual microscope fields.

Pulse-chase of newly synthesized RNA. HeLa cells were transfected with plasmids expressing GFP, (G_4C_2)₈ or ($G_4C_2)₅₈ that had been subcloned into pcDNA 3.1 using Fugene HD (Promega). After 48 h incubation, cells were treated with 1 mM 5-ethynyl uridine (EU) for 1 h. RNA was then visualized at the indicated time points using the Click-iT detection kit (Invitrogen) as recommended by the manufacturer. To visualize GFP signal, cells were co-stained using monoclonal anti-GFP (SC-9996, Santa Cruz Biotechnology) as a primary antibody.$

iPSC lines. iPSC lines from 2 control subjects and 4 G_4C_2 repeat expansion carriers have been previously characterized^{23–25,35} (see Supplementary Table 4). Generation of integration-free induced pluripotent stem cells (iPSCs) from fibroblasts of one G_4C_2 expansion carrier and one control subject, using episomal plasmids with the reprogramming factors Oct4, Sox2, Klf4, Lin28 and L-Myc, was performed as described previously³⁶, which was approved by the University of Massachusetts Medical School Institutional Biosafety Committee. The use of human fibroblasts was approved by the UCSF Institutional Review Board and informed consent was obtained from all subjects. After reprogramming, characterization of iPSCs shows that cells have normal karyotype, express pluripotent markers and have the capacity to differentiate into cells of the three germ layers as described by Almeida *et al.*^{23,35}. All the iPSC lines used here were tested regularly with no mycoplasma contamination.

Cortical neuronal cultures. Four iPSC lines from three control subjects and six iPSC lines from five *C9orf72* carriers were differentiated to cortical neurons as described earlier^{23,35} (see also Supplementary Table 5). Neuronal cultures were aged for 8 weeks before fixation for RNA-FISH.

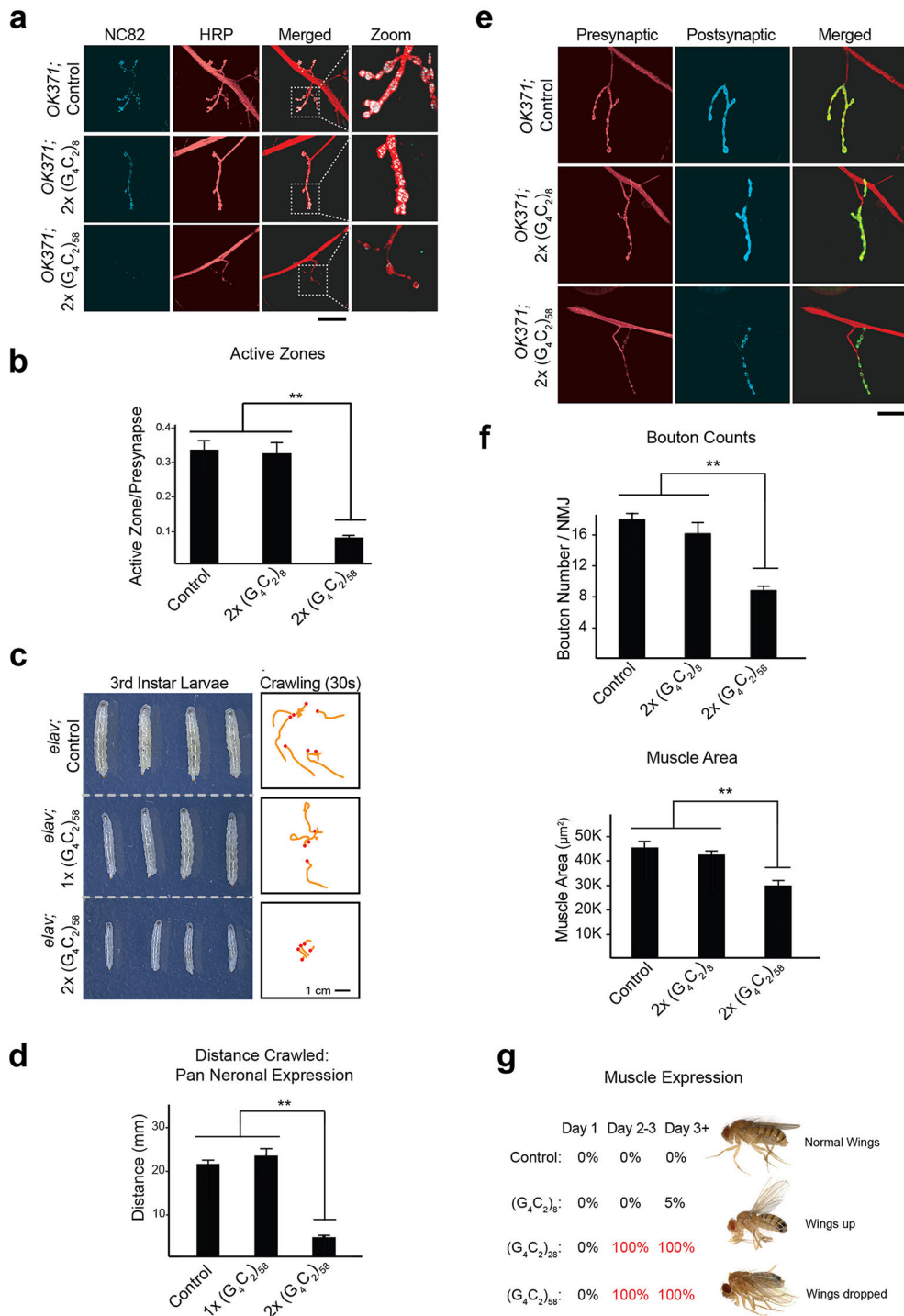
Immunostaining for iPSCs and iPSC-derived neurons. For immunostaining, cells were fixed in 4% paraformaldehyde for 15 min and then permeabilized with 0.3% Triton X-100 for 5 min. Cells were blocked with 5% bovine serum albumin for 30 min; cells were incubated with primary antibodies overnight at 4 °C. The following primary antibodies were used: mouse anti-OCT4 1:100 (Santa Cruz Biotechnology), goat anti-NANOG 1:100 (R&D Systems), mouse anti-SSEA4 1:100 (Abcam), rabbit anti-desmin 1:100 (Thermo Scientific), mouse anti-βIII-tubulin 1:500 (Promega), mouse anti-α-fetoprotein 1:200 (R&D Systems), mouse anti-MAP2 1:500 (Sigma), goat anti-ChAT 1:200 (EMD Millipore), rabbit anti-VGLUT1 1:500 (Synaptic Systems). After incubation with primary antibodies, cells were washed with PBS three times and incubated with Alexa Fluor secondary

antibodies 1:500 (Invitrogen) for 1 h at room temperature followed by counter staining with DAPI.

Fluorescence *in situ* hybridization. FISH in iPSCs derived from control and *C9orf72* expansion carrier using a Cy3-conjugated (GGCCCC)₄ probe was performed as described previously²³.

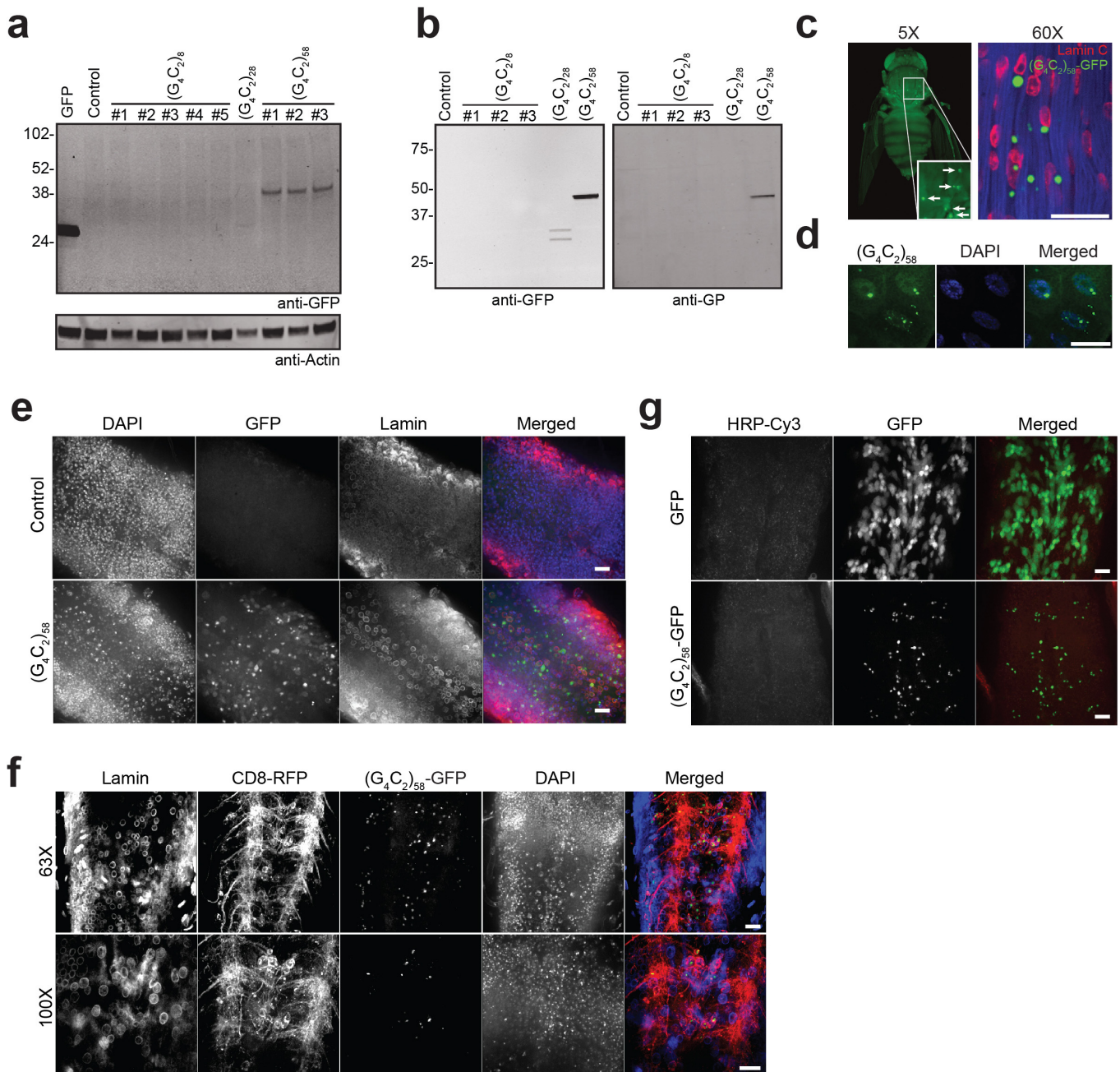
Quantification. For salivary glands, 30 cells from at least 5 individual salivary glands for each genotype were used for quantitative analysis. ImageJ software was used to measure the ratio of nuclear to cytoplasmic RNA by comparing the total RNA measured in the nucleus to that of the cytoplasm. DAPI was used to mark the nucleus. For cortical neurons, the ratio of nuclear to cytoplasmic RNA density was calculated by measuring the density of the RNA signal in both the nucleus and cell body using ImageJ software. DAPI was used to mark the nuclear boundary of neurons. For cortical neurons, at least 15 neurons were measured from each individual patient. Error bars for all quantification are standard error.

30. Zhang, Y. J. *et al.* Aggregation-prone c9FTD/ALS poly(GA) RAN-translated proteins cause neurotoxicity by inducing ER stress. *Acta Neuropathol.* **128**, 505–524 (2014).
31. Kim, N. C. *et al.* VCP is essential for mitochondrial quality control by PINK1/Parkin and this function is impaired by VCP mutations. *Neuron* **78**, 65–80 (2013).
32. Smith, R. & Taylor, J. P. Dissection and imaging of active zones in the *Drosophila* neuromuscular junction. *J. Vis. Exp.* (2011).
33. Gendron, T. F. *et al.* Antisense transcripts of the expanded C9ORF72 hexanucleotide repeat form nuclear RNA foci and undergo repeat-associated non-ATG translation in c9FTD/ALS. *Acta Neuropathol.* **126**, 829–844 (2013).
34. Ash, P. E. *et al.* Unconventional translation of C9ORF72 GGGGCC expansion generates insoluble polypeptides specific to c9FTD/ALS. *Neuron* **77**, 639–646 (2013).
35. Almeida, S. *et al.* Induced pluripotent stem cell models of progranulin-deficient frontotemporal dementia uncover specific reversible neuronal defects. *Cell Rep* **2**, 789–798 (2012).
36. Okita, K. *et al.* A more efficient method to generate integration-free human iPS cells. *Nature Methods* **8**, 409–412 (2011).



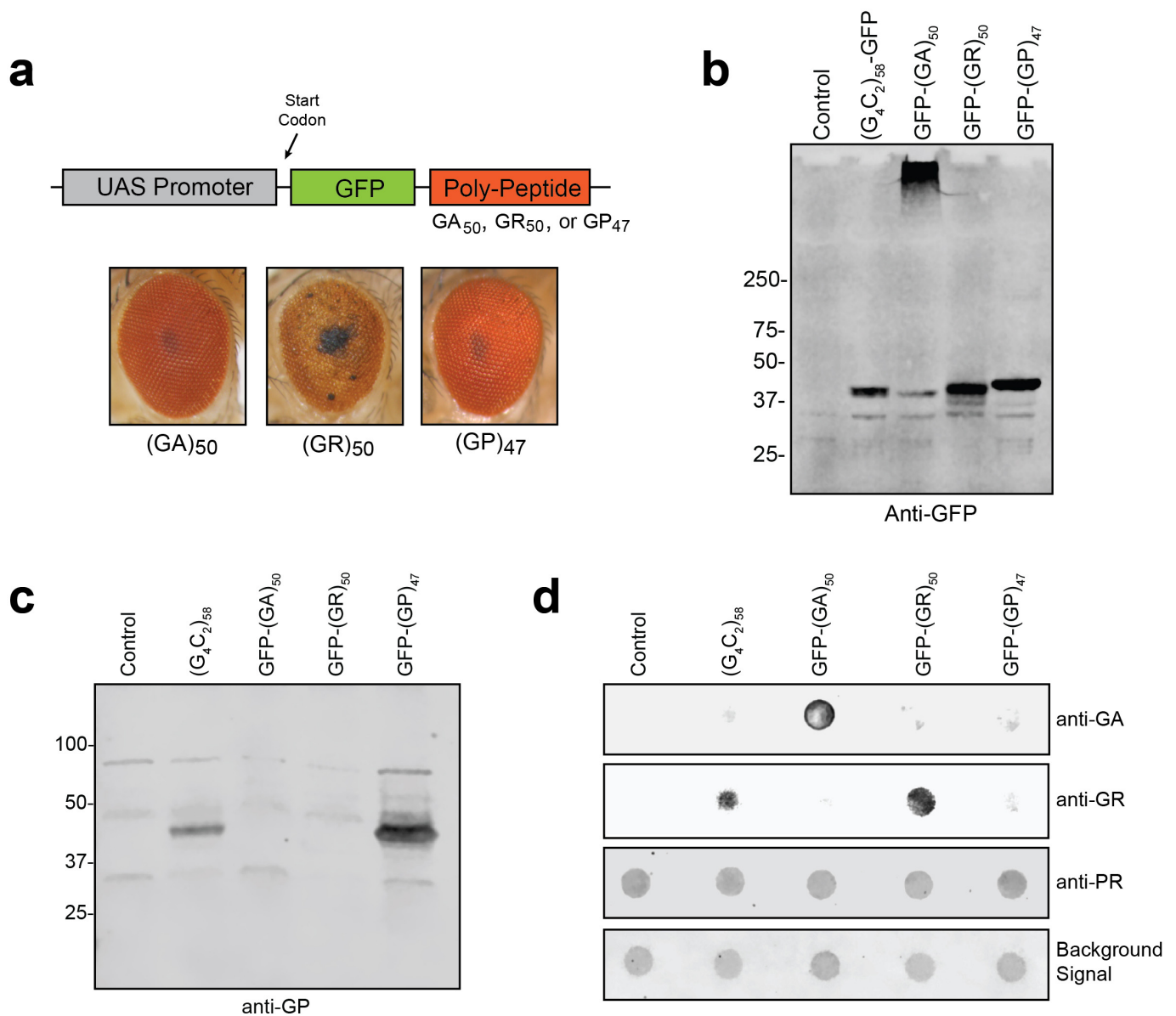
Extended Data Figure 1 | Expression of G₄C₂ repeats induces length-dependent phenotypes in *Drosophila*. **a**, Expression of (G₄C₂)₅₈ in *Drosophila* motor neurons using the OK371-GAL4 driver leads to a significant reduction in active zones as immunostained by the anti-Bruchpilot antibody NC82 and anti-HRP. Scale bar, 50 μm. **b**, Quantification of active zones ($n = 6$ individual larvae for control, 4 for 2× (G₄C₂)₈, and 6 for 2× (G₄C₂)₅₈). Values are mean ± s.e.m. ** $P < 0.01$, one-way ANOVA, Tukey's post hoc test. **c**, Pan neuronal expression of (G₄C₂)₅₈ repeats induces dosage-dependent decrease in larval size (left) and locomotor activity measured in 30 s (right) when (G₄C₂)₅₈ is expressed in all neurons using the *elav*-GAL4 driver. **d**, Quantification of the distance travelled by third instar larvae reveals expressing two copies of (G₄C₂)₅₈ results in a significant deficit in locomotor activity. Values are mean ± s.e.m. ($n = 7$ individual larvae for control, 4 for 1× (G₄C₂)₅₈, and 5 for 2' (G₄C₂)₅₈). ** $P < 0.01$, one-way ANOVA, Tukey's post

hoc test. **e**, Pan neuronal expression of (G₄C₂)₅₈ repeats in *Drosophila* neurons using the *elav*-GAL4 driver leads to a significant reduction in the bouton number. Bouton number was quantified by examining the presynaptic (anti-HRP) and postsynaptic (anti-DL1) markers (left). Scale bar, 50 μm. **f**, Quantification of bouton number (left) and muscle size (right) reveal that both are significantly reduced in *Drosophila* larvae expressing (G₄C₂)₅₈ repeats. Values are mean ± s.e.m. ($n = 6$ individual larvae for control, 5 for 2× (G₄C₂)₈, and 6 for 2× (G₄C₂)₅₈), ** $P < 0.01$, one-way ANOVA, Tukey's post hoc test. **g**, Expression of (G₄C₂)₂₈ and (G₄C₂)₅₈ but not (G₄C₂)₈ in the muscle using the *MHC*-GAL4 driver leads to loss of wing control in adult flies ($n = 30$ individual *Drosophila* for control, 21 for (G₄C₂)₈, 50 for (G₄C₂)₂₈, and 24 for (G₄C₂)₅₈). This phenotype was assessed by examining the permanent wing posture of live adult flies.



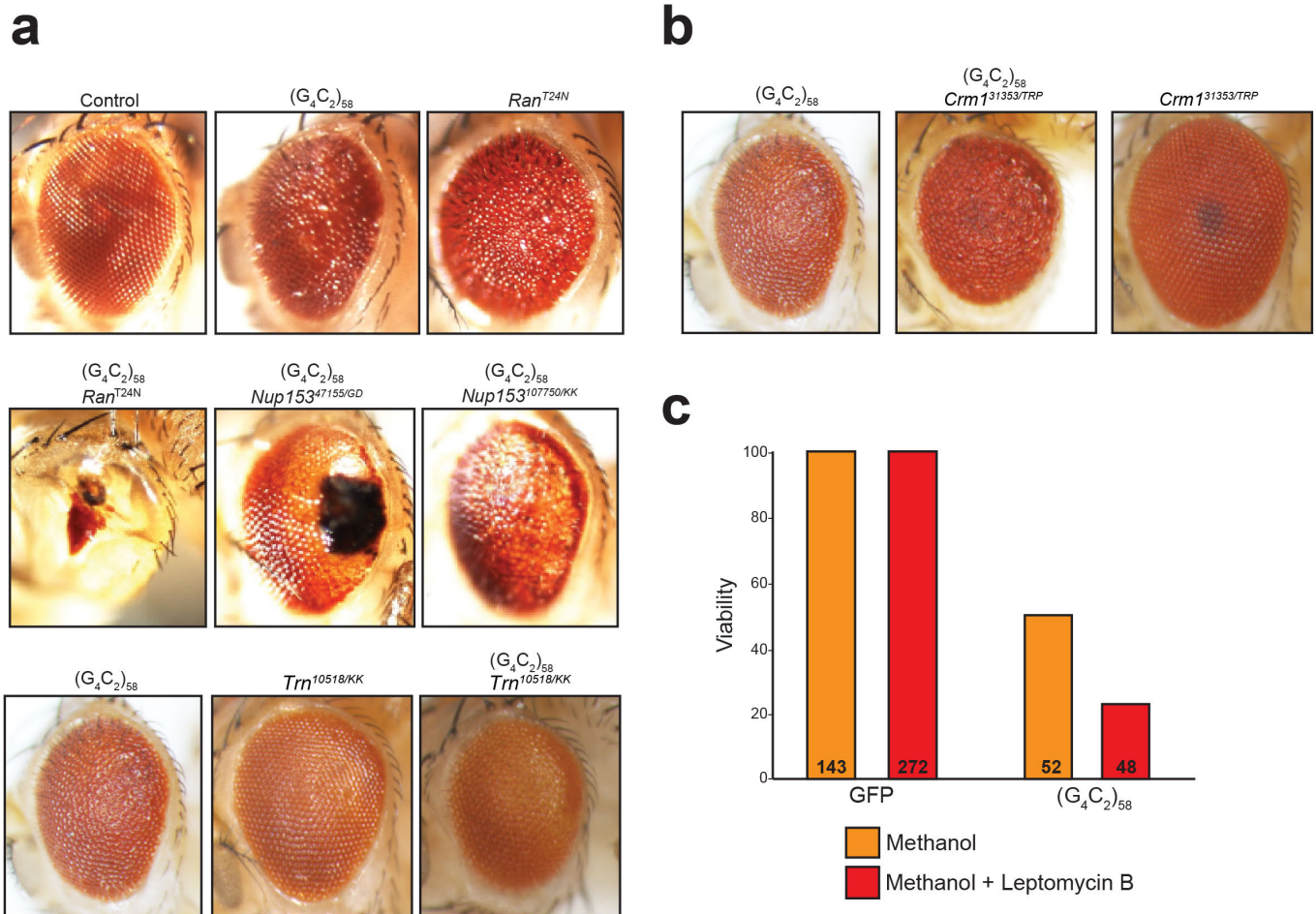
Extended Data Figure 2 | RAN translation is observed in *Drosophila* expressing G₄C₂ repeats. **a**, Western blot revealing translation of RAN poly-dipeptides in flies expressing (G₄C₂)₅₈ in the eye. RAN poly-dipeptides were not found in flies expressing (G₄C₂)₈ or control flies. There was minimal expression of GFP-positive product observed in flies expressing (G₄C₂)₂₈. GFP-expressing flies (lane 1) were used as a positive control for the anti-GFP antibody. **b**, Western blot showing production of RAN product when (G₄C₂)₂₈ and (G₄C₂)₅₈ but not (G₄C₂)₈ repeats are expressed in the muscle. RAN products were visualized with anti-GFP antibody (left) and anti-poly(GP) antibody (right). **c**, The RAN product poly-GP-GFP from flies expressing

(G₄C₂)₅₈ in the muscle form large visible inclusions as visualized under light sheet fluorescent microscopy (left) and by confocal microscopy (right). Scale bar, 50 μm. **d**, Expression of (G₄C₂)₅₈ in the salivary gland cells results in the formation of large nuclear inclusions and smaller cytoplasmic inclusions. Scale bar, 50 μm. **e**, **f**, Expression of (G₄C₂)₅₈ in the ventral ganglion by *OK371* driver results in the formation of nuclear and cytoplasmic inclusions, whereas GFP shows diffused nuclear and cytoplasmic localization. Lamin staining shows nuclear membrane, and CD8-RFP shows plasma membrane. Scale bar, 10 μm. **g**, Expression of (G₄C₂)₅₈ in pan neuronal cells by *elav* driver results in the nuclear and cytoplasmic inclusions. Scale bar, 10 μm.



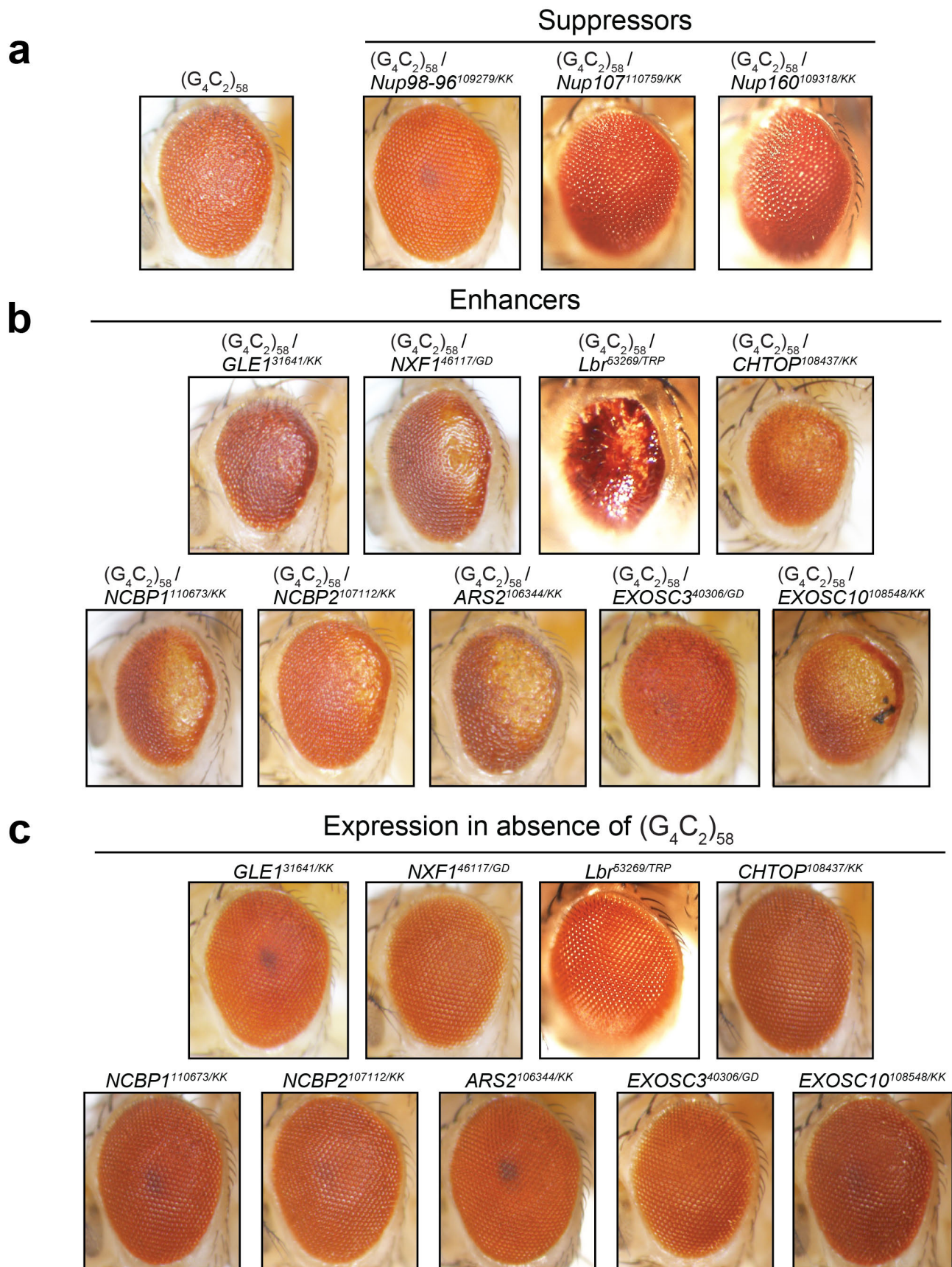
Extended Data Figure 3 | Ectopic expression of poly(GR) but not poly(GA) or poly(GP) peptides are toxic in *Drosophila*. **a**, Transgenic *Drosophila* were generated that express ATG-driven poly(GA), poly(GR) and poly(GP) peptides with an N-terminal GFP tag (top). Expression of GFP-(GA)₅₀ and GFP-(GP)₄₇ were non-toxic when expressed in the eye with *GMR-GAL4* whereas GFP-(GR)₅₀ expression resulted in >95% lethality with surviving adults having severely degenerated eyes (bottom). **b**, Western blot showing the expression of (G₄C₂)₅₈, GFP-(GA)₅₀, GFP-(GR)₅₀ and GFP-(GP)₄₇ as

visualized in muscle by anti-GFP antibody. **c**, Western blot showing the expression of poly(GP) in muscle of flies expressing (G₄C₂)₅₈ and GFP-(GP)₄₇ but not GFP-(GA)₅₀, GFP-(GR)₅₀ and control flies as visualized by anti-GP antibody. **d**, Dot blot analysis of RAN peptides in muscle revealing expression of poly(GA) only in GFP-(GA)₅₀ flies, expression of poly(GR) in (G₄C₂)₅₈ and GFP-(GR)₅₀ flies. As expected, anti-sense DPR poly(PR) was not found in any of the lysates. The background protein signal was used as a loading control.



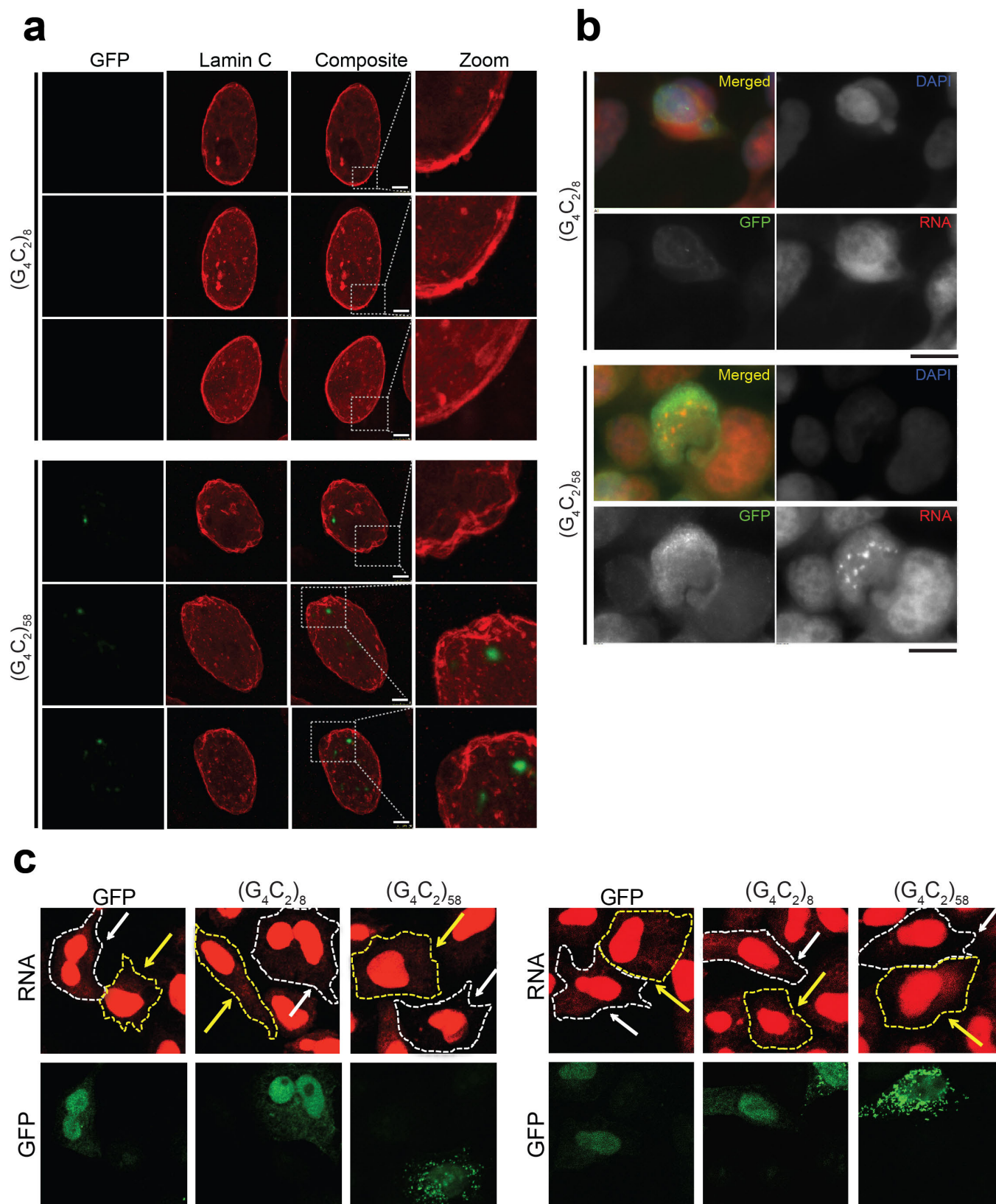
Extended Data Figure 4 | Nuclear import and export is altered by (G₄C₂)₅₈ expression. **a**, A threonine to asparagine substitution at residue 24 in the Ran protein abolishes the affinity for GTP and reduces its affinity for GDP. Hence, the Ran^{T24N} is always in either a nucleotide-free state or in its inactive, GDP-bound state, and acts as dominant negative. RAN^{T24N} expression driven by *GMR-GAL4* causes a mild eye phenotype when expressed in the absence of (G₄C₂)₅₈ (upper row, right panel). The (G₄C₂)₅₈ rough eye phenotype is strongly enhanced by dominant-negative Ran^{T24N} expression (middle row, left panel). The (G₄C₂)₅₈ eye phenotype is strongly enhanced by knockdown of Nup153 by two independent RNAi lines (middle row, two right

panels). The (G₄C₂)₅₈ eye phenotype is also mildly enhanced by knockdown of transportin (Trn) (bottom row). **b**, Knockdown of Crm1 in flies expressing (G₄C₂)₅₈ induces a mild enhancement of the (G₄C₂)₅₈ eye phenotype (left versus middle). Crm1 knockdown in the absence of (G₄C₂)₅₈ repeats does not produce a rough eye phenotype (right). **c**, Expression of two copies of (G₄C₂)₅₈ in the *Drosophila* motor neurons leads to reduced viability (50%). Chemical inhibition of Crm1 with Leptomycin B (500 nM) enhances (G₄C₂)₅₈ toxicity resulting in reduced viability (23%). Leptomycin B does not impede viability (100%) in *Drosophila* expressing GFP. *n* is displayed on the graph and represents the individual pupae from two separate experiments.



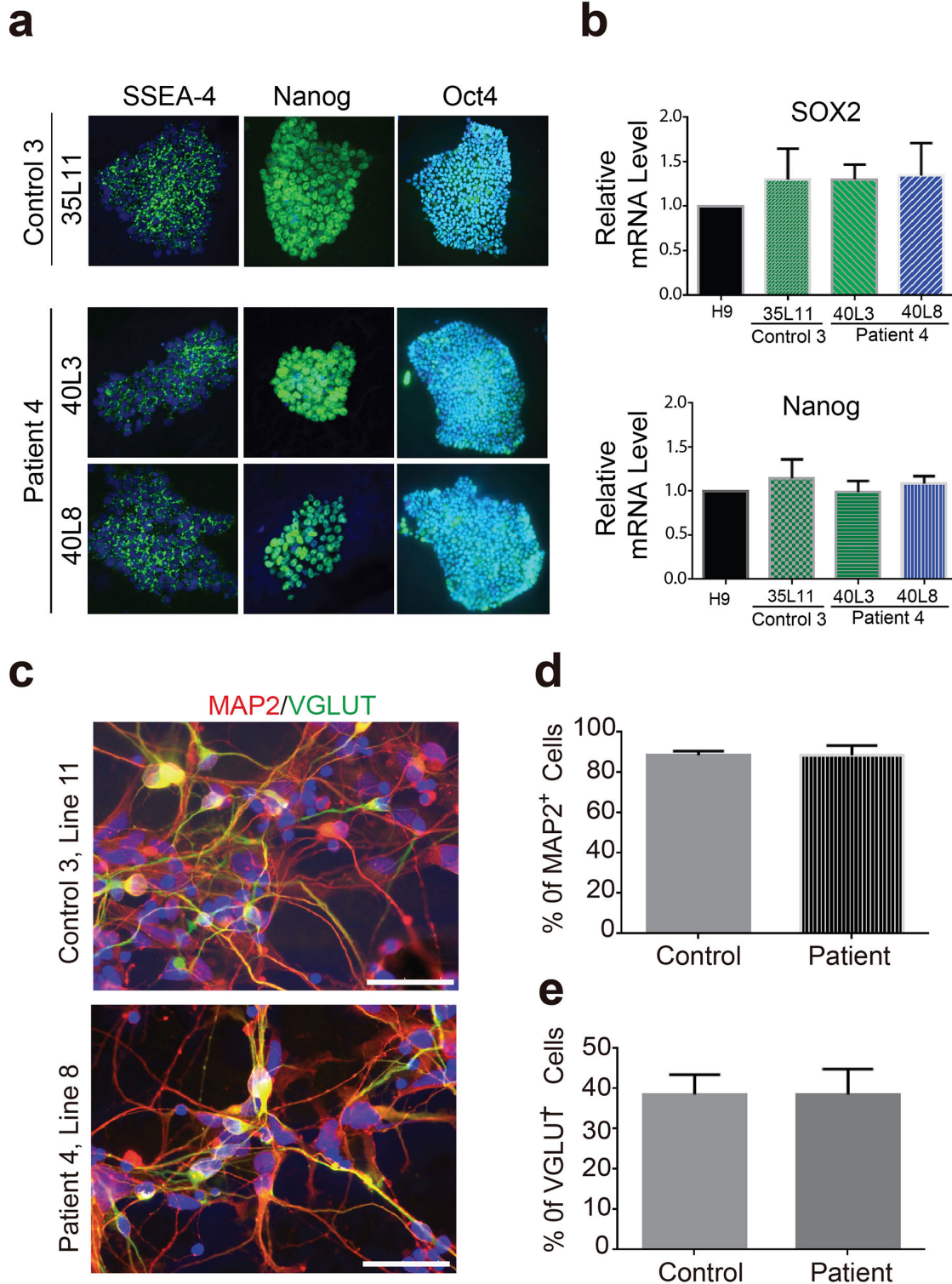
Extended Data Figure 5 | Phenotypes of additional suppressors and enhancers of (G₄C₂)₅₈. **a**, Phenotypes demonstrating suppression of the (G₄C₂)₅₈ rough eye phenotype by RNAi knockdown of identified genes. **b**, Phenotypes demonstrating enhancement of the (G₄C₂)₅₈ rough eye

phenotype by RNAi knockdown of identified genes. **c**, Knockdown of identified modifier genes shows little or no phenotype in the absence of G₄C₂ repeat expression.



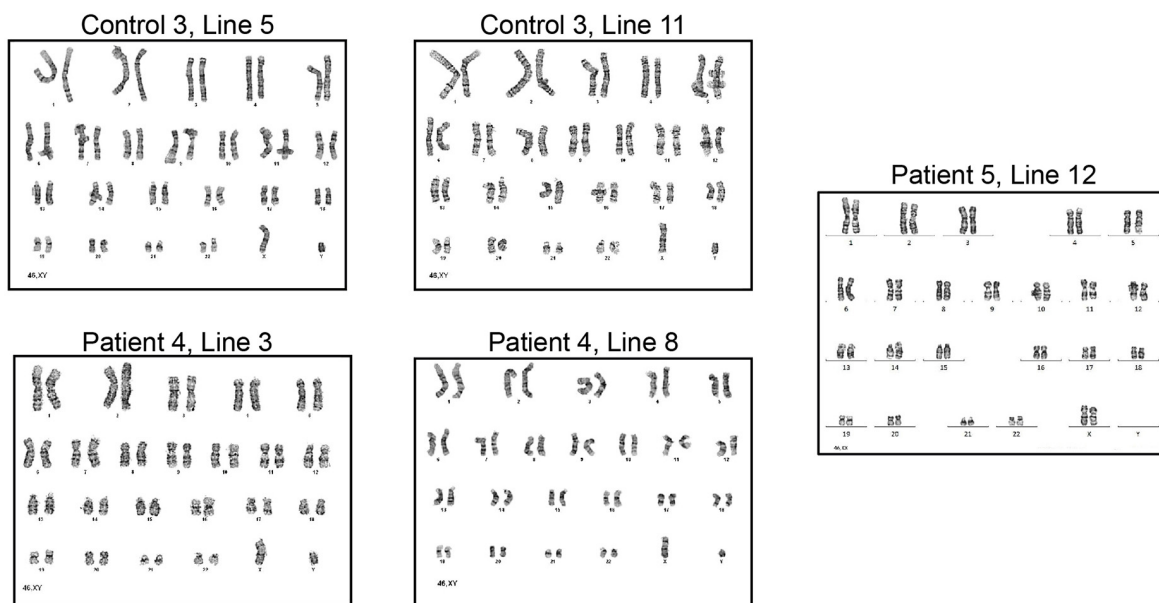
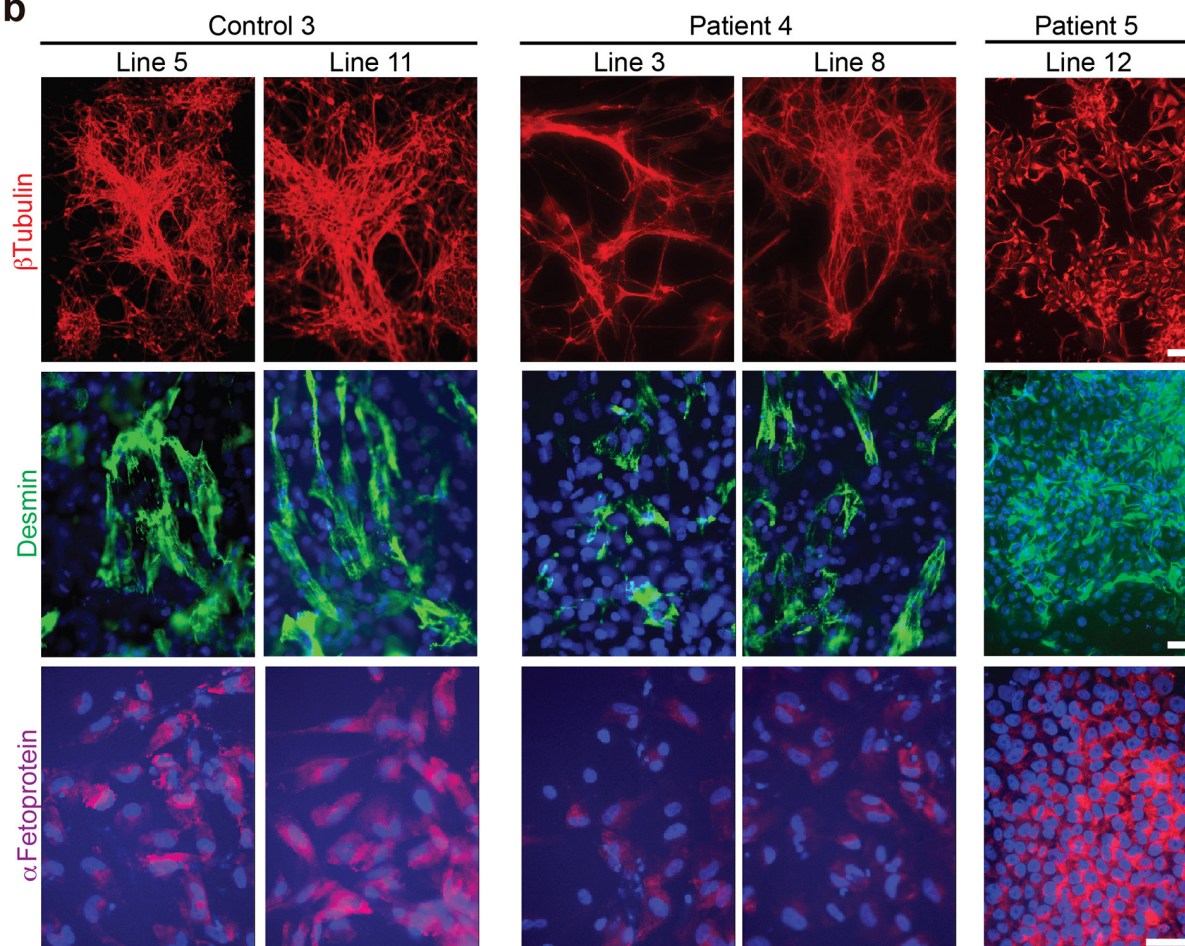
Extended Data Figure 6 | Impairment of nucleocytoplasmic shuttling in *Drosophila* and cultured human cell lines. **a**, $(G_4C_2)_{58}$ expression driven by *Fkh-GAL4* causes an abnormal nuclear envelope as shown by Lamin C staining (bottom) in comparison to $(G_4C_2)_8$ (top). Scale bar, 10 μ m. **b**, Transfection of 293T cells with $(G_4C_2)_{58}$ (bottom) but not $(G_4C_2)_8$ (top) leads to an increase in

nuclear RNA puncta as visualized with a total RNA-FISH probe. Non-transfected cells (absence of GFP signal) do not show an increase in nuclear RNA in either $(G_4C_2)_8$ or $(G_4C_2)_{58}$ transfected cells. Scale bar, 25 μ m. **c**, Enlarged images showing slowed accumulation of newly synthesized RNA in the cytoplasm of HeLa cells expressing $(G_4C_2)_{58}$. Scale bar, 25 μ m.



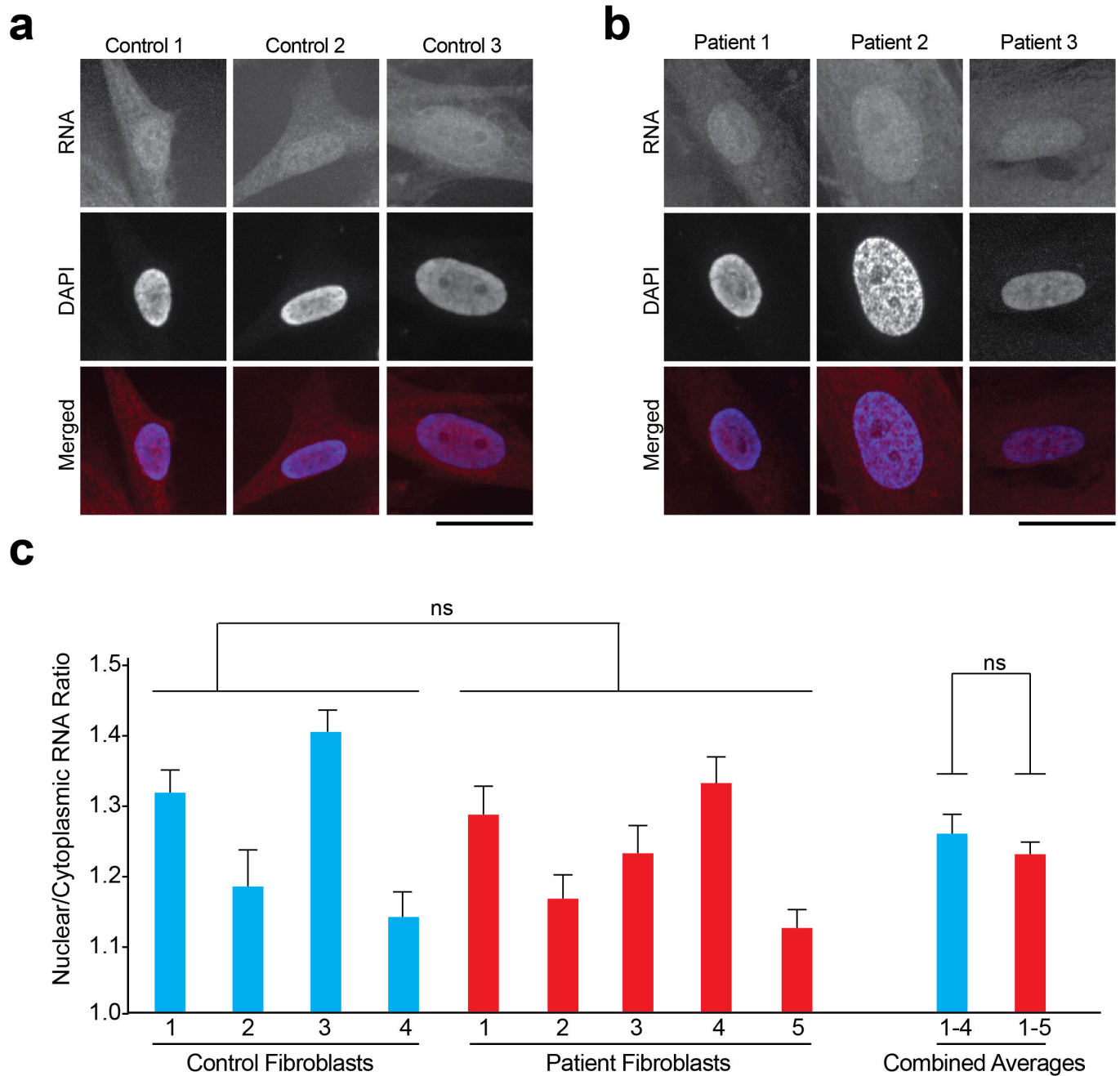
Extended Data Figure 7 | Characterization of newly generated integration-free iPSC lines. **a**, iPSC lines from a control subject (line 11) and a G_4C_2 repeat expansion carrier (line 3 and line 8) express pluripotent markers SSEA-4, Nanog and Oct-4. Scale bar, 50 μ m. **b**, qRT-PCR analysis of expression levels of pluripotent stem-cell markers SOX2 and Nanog in these iPSC lines showing no statistical differences between these lines and human embryonic stem-cell line H9. Relative mRNA levels are quantified from 3 independent

experiments. **c**, After differentiation into cortical neurons about 90% of cells in these cultures are MAP2-positive neurons. Scale bar, 50 μ m. **d**, Quantification of average percentage of MAP2-positive neurons and there is no difference between control and *C9orf72* cultures. Average percentages were quantified from 3 independent experiments. **e**, Quantification of average percentage of VGLUT-positive excitatory neurons among all neurons; there is no difference between control and *C9orf72* cultures. $n = 3$ independent experiments.

a**b**

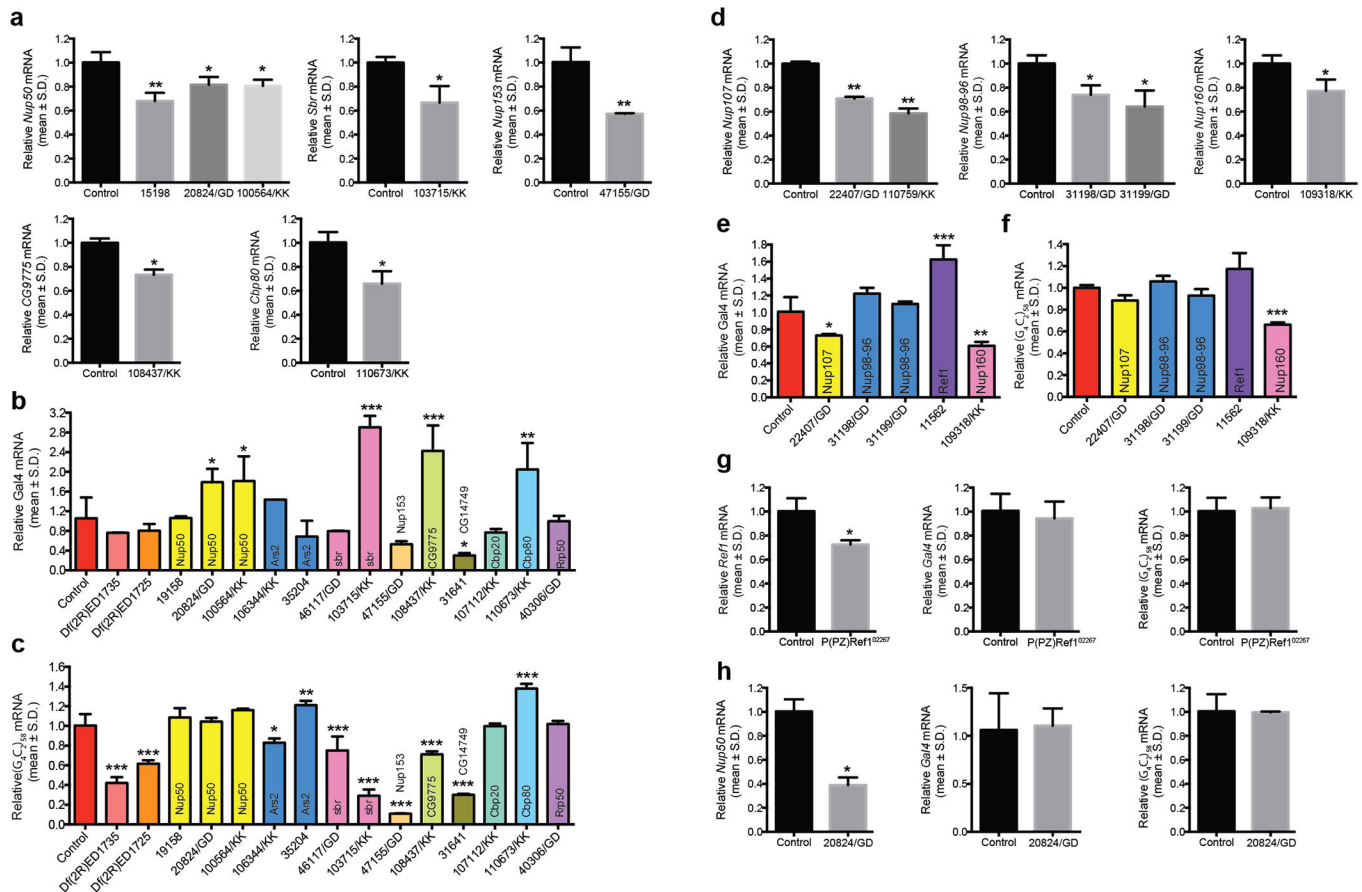
Extended Data Figure 8 | Karyotyping analysis and pluripotency of newly generated iPSC lines. **a**, G-band staining showing a normal karyotype for all the lines analysed. **b**, After *in vitro* spontaneous differentiation of control and C9 carrier iPSC lines, cells were stained for α -fetoprotein (endoderm),

desmin (mesoderm), β III-tubulin (ectoderm) and Hoechst (nuclei). All lines showed differentiation towards derivatives of three germ layers. Scale bars, 20 μ m.



Extended Data Figure 9 | Accumulation of nuclear RNA is not seen in fibroblasts derived from patients with G_4C_2 repeat expansion. **a, b,** Total cellular RNA was measured by FISH in fibroblasts derived from 4 control (a) subjects or 5 patients (b) with G_4C_2 repeat expansion. Scale bar, 25 μ m.

c, Quantification shows no statistical difference in the observed nuclear to cytoplasmic RNA ratio in patient versus control fibroblasts. $n = 16$ individual cells analysed for each line.



Extended Data Figure 10 | qRT-PCR analysis. **a–f**, qRT-PCR analysis demonstrating knockdown of selected modifiers in *Drosophila* eyes. mRNA levels of selected modifier (**a**, **d**), *GAL4* (**b**, **e**) and $(G_4C_2)_{58}$ (**c**, **f**) assayed by qRT-PCR in progeny resulting from wild type (w1118), classical mutant allele or UAS RNAi lines of selected modifiers mated with either *GMR-GAL4* or *GMR-GAL4/Cy0*; UAS- G_4C_2 -58-GFP/TM6 to induce knockdown of the selected gene. RNA was obtained from whole *Drosophila* head lysates. Gene expression levels are mean \pm s.d. from $n = 3$ independent experiments,

* $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$ by one-way ANOVA, Tukey's post hoc test. **g**, **h**, qRT-PCR analysis demonstrating knockdown of Ref1 and Nup50 in salivary gland. mRNA levels of selected modifier (left), *GAL4* (middle) and $(G_4C_2)_{58}$ (right) assayed by qRT-PCR in progeny resulting from either P(PZ)Ref1⁰²²⁶⁷ (**g**) or Nup50^{20824/GD} (**h**) mated with *Fkh-GAL4*; UAS- G_4C_2 -58-GFP/TM6. RNA was obtained from salivary gland lysates. Gene expression levels are mean \pm s.d., $n = 3$ independent experiments. * $P < 0.05$ by Student's *t*-test.

Orientation-specific joining of AID-initiated DNA breaks promotes antibody class switching

Junchao Dong^{1*}, Rohit A. Panchakshari^{1*}, Tingting Zhang^{1*†}, Yu Zhang^{1†}, Jiazhi Hu¹, Sabrina A. Volpi², Robin M. Meyers¹, Yu-Jui Ho^{1†}, Zhou Du¹, Davide F. Robbiani³, Feilong Meng¹, Monica Gostissa^{1†}, Michel C. Nussenzweig³, John P. Manis² & Frederick W. Alt¹

During B-cell development, RAG endonuclease cleaves immunoglobulin heavy chain (IgH) V, D, and J gene segments and orchestrates their fusion as deletional events that assemble a V(D)J exon in the same transcriptional orientation as adjacent C_μ constant region exons^{1,2}. In mice, six additional sets of constant region exons (C_Hs) lie 100–200 kilobases downstream in the same transcriptional orientation as V(D)J and C_μ exons². Long repetitive switch (S) regions precede C_μ and downstream C_Hs. In mature B cells, class switch recombination (CSR) generates different antibody classes by replacing C_μ with a downstream C_H (ref. 2). Activation-induced cytidine deaminase (AID) initiates CSR by promoting deamination lesions within S_μ and a downstream acceptor S region^{2,3}; these lesions are converted into DNA double-strand breaks (DSBs) by general DNA repair factors³. Productive CSR must occur in a deletional orientation by joining the upstream end of an S_μ DSB to the downstream end of an acceptor S-region DSB. However, the relative frequency of deletional to inversional CSR junctions has not been measured. Thus, whether orientation-specific joining is a programmed mechanistic feature of CSR as it is for V(D)J recombination and, if so, how this is achieved is unknown. To address this question, we adapt high-throughput genome-wide translocation sequencing⁴ into a highly sensitive DSB end-joining assay and apply it to endogenous AID-initiated S-region DSBs in mouse B cells. We show that CSR is programmed to occur in a productive deletional orientation and does so via an unprecedented mechanism that involves *in cis* IgH organizational features in combination with frequent S-region DSBs initiated by AID. We further implicate ATM-dependent DSB-response factors in enforcing this mechanism and provide an explanation of why CSR is so reliant on the 53BP1 DSB-response factor.

Most chromosomal DSB ends join to ends of separate DSBs genome-wide without orientation (end) specificity^{4,5}. Similarly, non-productive 'inversional' CSR joins have been found in transformed B cells^{6–9}, suggesting CSR also may not be orientation-specific¹⁰ (Fig. 1a). To address this possibility, we employed digestion–circularization PCR (DC–PCR, Extended Data Fig. 1a) to identify the orientation of CSR joins between S_μ and S_{γ1} in purified mouse B cells stimulated with anti-CD40 plus IL4 to activate AID-targeting to S_{γ1} and S_ε, and class-switching to IgG1 (and IgE). Most S_μ to S_{γ1} junctions identified by this semi-quantitative approach were deletional (Extended Data Fig. 1b).

To confirm DC–PCR findings and analyse potential mechanisms, we used high-throughput genome-wide translocation sequencing (HTGTS), an unbiased genome-wide approach that identifies 'prey' DSB junctions to a fixed 'bait' DSB with nucleotide resolution^{4,5} (Extended Data Fig. 1c). We refer to broken ends of bait IgH DSBs

as 5'- and 3'-broken ends; specific primers allow use of each as bait⁴ (Fig. 1b, c). Prey junctions are denoted + if prey is read from the junction in a centromere-to-telomere direction and – if in the opposite direction⁴ (Fig. 1b, c). The + and – outcomes for intra-chromosomal joining of broken ends of different DSBs on the same chromosome include rejoining of a DSB subsequent to resection, or joining the broken ends of two separate DSBs to form intra-chromosomal inversions, deletions, or excision circles^{4,5} (Fig. 1b, c). To assess the relative frequency at which non-AID-initiated IgH DSBs join in deletional versus inversional orientation, we expressed I-SceI endonuclease in anti-CD40/IL4-activated AID-deficient B cells in which I-SceI targets were inserted upstream of S_μ and downstream of S_{γ1} (Igh^{I-96k} allele¹¹; Extended Data Fig. 1d, e), or in AID-sufficient B cells in which S_{γ1} and S_μ were replaced with I-SceI targets (ΔS_μ^{2×1}/ΔS_{γ1}^{2×1} allele¹²; Fig. 1d and Extended Data Fig. 1f). HTGTS with primers that captured junctions involving 3'- or 5'-broken ends of I-SceI bait DSBs in the S_{γ1} locale revealed that a major class of recovered junctions were re-joins of bait DSBs following resection (Fig. 1d and Extended Data Fig. 1d–f). A second major class of bait junctions in the S_{γ1} locale involved intact or resected 3'- or 5'-broken ends of I-SceI-generated DSB in the S_μ locale, which comprised relatively similar numbers of deletional (+) and inversional (–) junctions for bait 3'-broken ends (Fig. 1d and Extended Data Fig. 1d) and similar numbers of excision circle (–) versus inversional (+) junctions for bait 5'-broken ends (Extended Data Fig. 1e, f). As expected⁴, bait 3'- and 5'-broken ends from the S_{γ1} locale recovered similar levels of + and – junctions genome-wide (Extended Data Fig. 2a–d). We conclude that joining between two I-SceI DSBs in different IgH S-region locations in CSR-activated B cells lacks any notable preference for or against inversional versus deletional joins.

In AID-deficient Igh^{I-96k} B cells, I-SceI 5'- and 3'-broken end baits downstream of S_{γ1} did not capture IgH DSB hotspots beyond I-SceI-generated broken ends upstream of S_μ (Extended Data Fig. 1d, e). In contrast, I-SceI 5'- and 3'-broken ends from the ΔS_μ^{2×1}/ΔS_{γ1}^{2×1} allele in AID-sufficient B cells joined frequently to AID-initiated S_ε DSBs 60 kilobases (kb) downstream (Fig. 1d and Extended Data Fig. 1f), with the majority (~80%) of 3' and 5' ΔS_μ^{2×1}/ΔS_{γ1}^{2×1} broken end joins distributed across the 4-kb S_ε in orientations that generate, respectively, excision circles (Fig. 1d) or deletions (Extended Data Fig. 1f). We also performed HTGTS on activated, I-SceI-expressing B cells in which only S_{γ1} was replaced by an I-SceI cassette (ΔS_{γ1}^{2×1} allele¹²; Fig. 1e). Beyond break-site junctions, major IgH hotspot regions of 3' ΔS_{γ1}^{2×1} broken ends were S_μ and S_ε (Fig. 1e and Extended Data Fig. 2j). Junctions occurred broadly across S_μ, with 80% in a deletional orientation; while 90% of S_ε junctions were in the reciprocal excision circle orientation (Fig. 1e; Extended Data Fig. 2j). CH12F3 B

¹Howard Hughes Medical Institute, Program in Cellular and Molecular Medicine, Boston Children's Hospital, and Department of Genetics, Harvard Medical School, Boston, Massachusetts 02115, USA.

²Boston Children's Hospital and Joint Program in Transfusion Medicine, Harvard Medical School, Boston, Massachusetts 02115, USA. ³Howard Hughes Medical Institute, Laboratory of Molecular Immunology, The Rockefeller University, New York, New York 10065, USA. [†]Present addresses: Eli Lilly and Company, Alexandria Center for Life Sciences, 450 East 29th Street, New York, New York 10016, USA (T.Z.); National Institute of Biological Sciences, Beijing 102206, Beijing, China (Y.Z.); Cold Spring Harbor Laboratory, Watson School of Biological Sciences, Cold Spring Harbor, New York 11724, USA (Y.-J.H.); 121Bio, 700 Main Street, Cambridge, Massachusetts 02139, USA (M.G.).

*These authors contributed equally to this work.

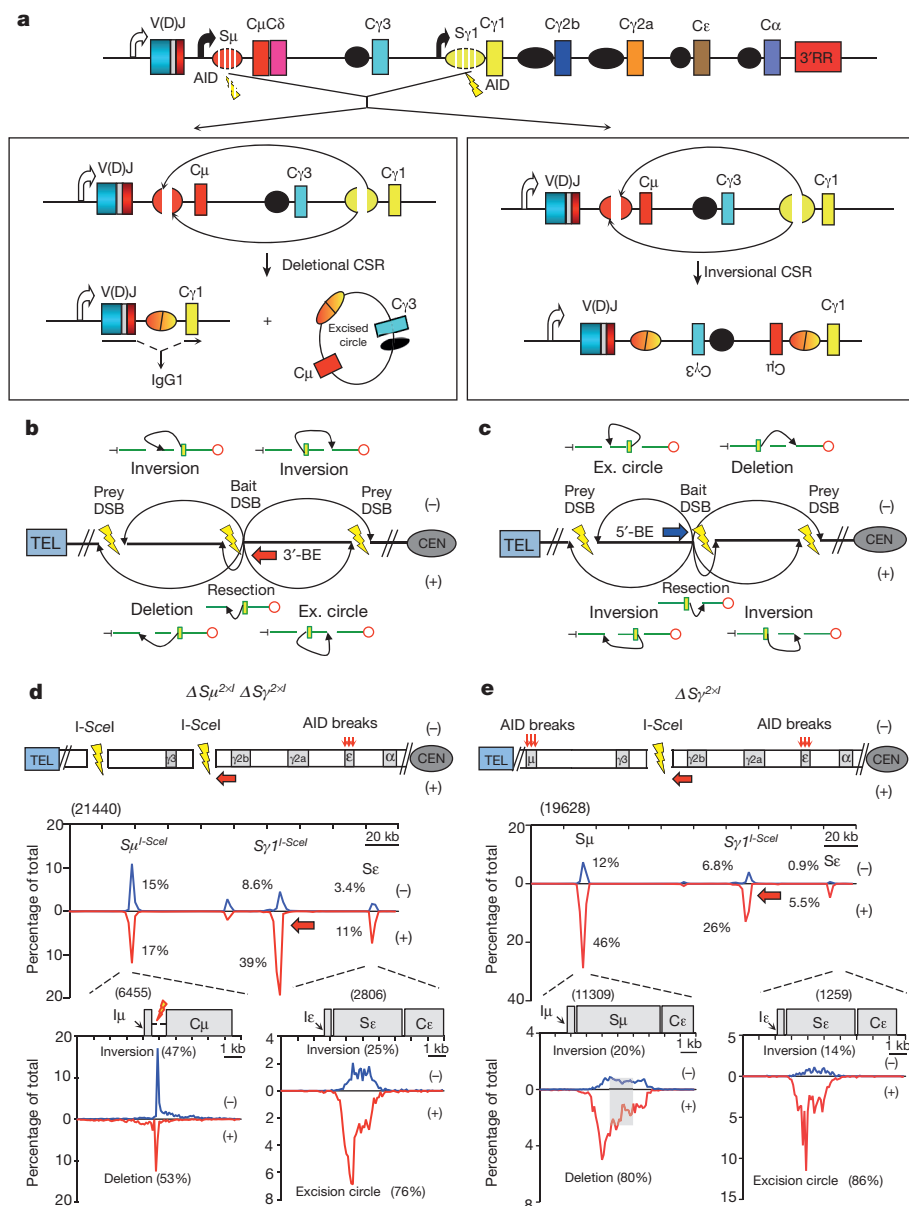


Figure 1 | S-region-dependent orientation-biased joining in CSR-stimulated B cells. **a**, Top, *Igh* C μ locus with AID targeting (lightning bolts) in S μ and S γ 1. Bottom, productive (left) CSR via deletional S μ to S γ 1 joins plus excision circles; and inversional (right) CSR. 3'RR, 3' regulatory region. **b, c**, Joining outcomes from 3' (red arrow) or 5' (blue arrow) bait broken ends (BEs; bolts) to prey broken ends⁵. CEN, centromere; TEL, telomere. **d**, Top, joining between I-SceI DSBs at S γ 1 and S μ in $\Delta S\mu^{2\times 1} \Delta S\gamma^{2\times 1}$ B cells¹². Middle, location of bait BE junctions to I-SceI-generated $\Delta S\mu^{2\times 1}$ or AID-initiated S ϵ prey broken ends in either + (red) or - (blue) orientation (five experiments) as a percentage of total *Igh* junctions (2-kb bins). Bottom, junction distribution to the 10-kb region encompassing I-SceI prey DSBs at S μ and AID-initiated prey DSBs at S ϵ . **e**, Results from $S\gamma^{12\times 1/+}$ B cells¹² plotted as in **d**. Grey box indicates prey junctions not assignable to single core-S μ sequence. Numbers in parenthesis denote total unique junctions. Statistical analyses in Extended Data Fig. 2h, j.

lymphoma cells in which S α was replaced with an I-SceI site had a similar orientation bias of S α I-SceI 3'-broken end joining to S μ DSBs (Extended Data Fig. 2n–q). Joining of the 5'-broken ends of $\Delta S\mu^{2\times 1}$ (on the $\Delta S\mu^{2\times 1}/\Delta S\gamma^{12\times 1}$ allele) to AID-initiated DSBs in S γ 3, S γ 2b and S γ 2a in lipopolysaccharide plus anti-IgD-dextran-activated B cells were similarly orientation-biased (Extended Data Fig. 3a–c). However, joining of the 5'-broken ends of $\Delta S\mu^{2\times 1}$ across an array of 28 \times I-SceI sites replacing S γ 1¹³ was not orientation-biased (Extended Data Fig. 3d, e). Together, these findings suggest that orientation-specific CSR joining requires an S-region sequence and/or unique aspects of S-region DSBs.

Mammalian S regions are G-rich on the non-template strand, giving AID-initiated 5' and 3' S-region broken ends a potential end-sequence bias. Also, when transcribed in the sense direction, S regions generate stable R-loops^{14,15}, which could differentially affect 5' and 3' S-region broken end structure. To test the potential roles of S regions in orientation-specific CSR, we used a Cas9/gRNA approach to invert S μ on the productive allele of CH12F3 B cells, which modestly reduced CSR (Extended Data Fig. 3f–h). We then assayed CH12F3 cells in which S α was replaced with an I-SceI site and S μ was in a normal or inverted orientation. These assays revealed that joins of I-SceI-generated 3'-broken ends at the S α locale to S μ DSBs were similarly biased for deletional junctions independent of S μ orientation (Fig. 2a–c).

Consistent with low-level *trans* CSR¹⁶, HTGTS libraries from activated $\Delta S\mu^{2\times 1}/\Delta S\gamma^{12\times 1}$ B cells contained numerous junctions from $\Delta S\gamma^{12\times 1}$ 3'-broken ends across the *trans* S μ ; which, in contrast to *cis* $\Delta S\gamma^{12\times 1}$ 3'-broken end S μ junctions, occurred in + and - orientations at a similar frequency (Fig. 2d). Likewise, bait 3'-broken ends from the $\Delta S\gamma^{12\times 1}$ *Igh* allele identified approximately equal numbers of (+) versus (-) junctions to AID off-target DSBs in *Il4ra* on chromosome 7 (Extended Data Fig. 2e). Finally, translocations between bait 5' I-SceI DSB broken ends in *c-myc*⁴ and prey AID-initiated S μ and S ϵ broken ends in CSR-activated B cells lacked orientation bias (Fig. 2e). We conclude that orientation-dependent CSR joining does not require orientation-associated features of S μ sequence, transcription, or transcripts. Moreover, AID-initiated DSBs *per se* are not sufficient to promote orientation specificity, as demonstrated by orientation-independence of DSB joining to them *in trans*. Thus, beyond S-region sequences and/or high frequency AID-initiated DSBs within them, aspects of *Igh* locus organization *in cis* must play a critical role in promoting orientation-dependent CSR joining.

We tested whether joining between two sets of endogenous AID-initiated S-region DSBs is orientation-dependent. Use of core S-region DSBs as HTGTS bait is confounded by their highly repetitive nature. Therefore, we used as bait a 150-base-pair (bp) sequence at the 5' end of S μ (5'S μ), which retains 14 of approximately 500 S μ AID-target

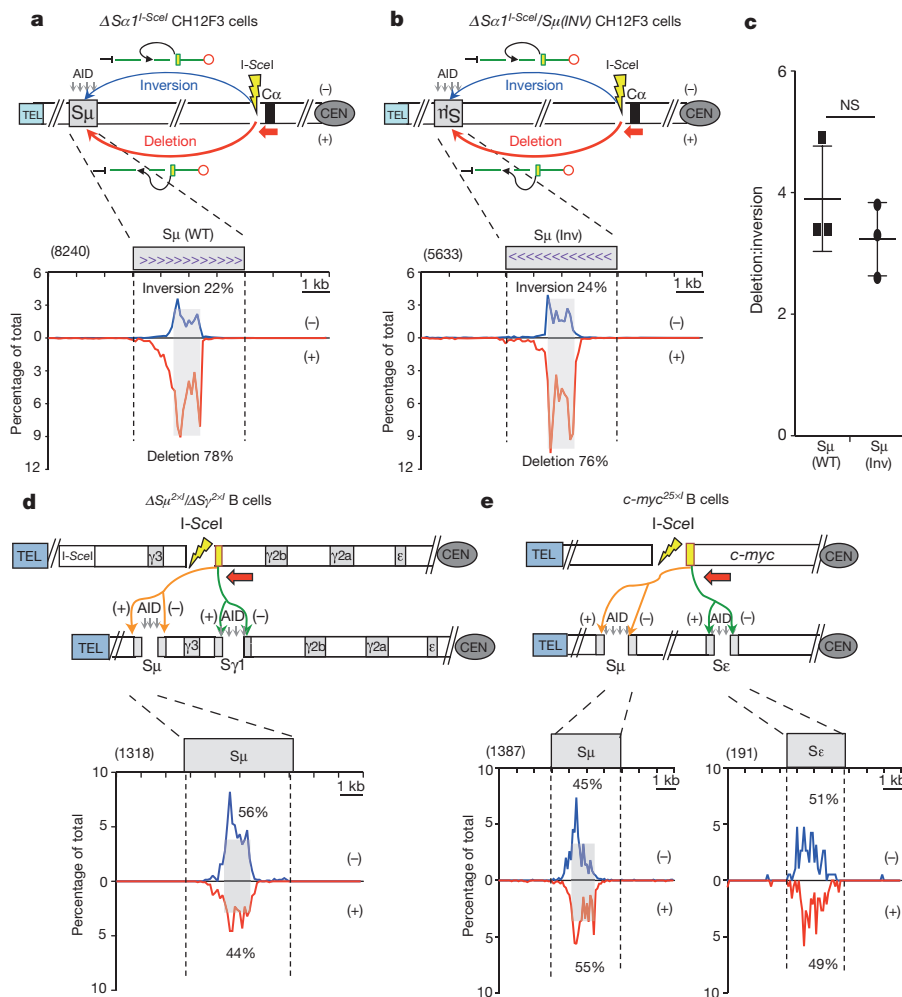


Figure 2 | S regions are not sufficient to promote orientation-biased CSR joining.

a, b, Top, HTGTS in CSR-activated CH12F3 cells for joining between 3'-broken ends of I-SceI DSB in the cassette replacing S α and AID-initiated S μ DSBs in which S μ is in the normal (**a**; WT) or inverted (**b**; Inv) orientation. Bottom, junctions from I-SceI DSB 3'-broken ends to normal (left) or inverted (right) S μ plotted as in Fig. 1d. **c**, Plot showing ratio (average \pm s.d.) of + (deletional) to - (inversional) joins in wild-type S μ (squares, $n = 3$) and S μ (Inv) (circles, $n = 3$) cells, with significance calculated by unpaired two-tailed t -test ($P = 0.307$). NS, not significant. **d**, Top, HTGTS junctions from $\Delta S\gamma I^{2\times I}$ 3'-broken ends to AID-induced S μ broken ends on *trans* chromosome in + and - orientations as in Fig. 1d. Bottom, distribution of junctions in 10-kb *trans* S μ ($n = 5$). **e**, Top, joining of bait I-SceI 5'-broken ends of $c-myc^{25\times I}$ cassette⁴ on chromosome 15 and AID-induced S μ and S ϵ Igh breaks in + and - orientations as in Fig. 1d. Bottom, linear distribution of junctions in S μ (left) or S ϵ (right) ($n = 4$). Statistical analysis in Extended Data Fig. 2k-m. Numbers in brackets in **d** and **e** denote biological replicates.

motifs (Fig. 3a, left panel). HTGTS of anti-CD40/IL4-stimulated B cells with the 5'S μ broken end primer revealed break-site junctions, as well as S $\gamma 1$ and S ϵ junctions (Fig. 3b, c). Consistent with AID-initiation, bait junctions were enriched at AID-targets within the 5'S μ bait (Fig. 3a, right panel). 5'S μ broken end junctions spread broadly over prey S regions, with up to 95% in a deletional orientation (Fig. 3c). For comparison, we tested a 150-bp 5' remnant of S μ (rS μ ; Extended Data Fig. 4a, left panel), retained when the rest of S μ was deleted¹⁷. B cells homozygous for rS μ have reduced IgG1 CSR but nearly normal IgE CSR¹⁸. HTGTS with either 5' rS μ or 3' rS μ broken end primers of anti-CD40/IL4- and lipopolysaccharide/anti-IgD-dextran-stimulated B cells, respectively, revealed junctions to S $\gamma 1$ and S ϵ and to S $\gamma 3$, S $\gamma 2b$, and S $\gamma 2a$ (Extended Data Fig. 4). 5' rS μ broken end junctions spread over target S regions, with >90% in a deletional orientation (Extended Data Fig. 4b, f); while >90% of 3' rS μ broken end junctions were in the complementary excision circle orientation (Extended Data Fig. 4c, g). Within the bait rS μ , junctions again were enriched at AID targets (Extended Data Fig. 4a). Consistent with IgH class-switching patterns, rS μ HTGTS junctions occurred more frequently to S ϵ than those from the 5'S μ bait in the context of full-length S μ (Extended Data Fig. 4b). Analyses of rS μ -mutant CH12F3 cells gave similar results (Extended Data Fig. 5a-c). Thus, AID-initiated S μ DSB joining to all downstream acceptor S regions is strongly biased towards the deletional orientation.

CSR DSBs generate a DSB response (DSBR) in which ATM activates histone H2AX and 53BP1 in chromatin flanking DSBs, thereby contributing to end-joining¹⁹⁻²¹. ATM or H2AX deficiency moderately reduces CSR (Extended Data Fig. 6a)^{2,19}. However, 53BP1 deficiency causes a more drastic reduction (Extended Data

Fig. 6a), suggesting specialized CSR roles^{2,19,22}, such as promoting S-region synapsis or protecting S-region DSBs from resection^{11,23-25}. To elucidate influences on orientation-specific CSR, we employed HTGTS to assay joining of AID-initiated 5'S μ broken ends to AID-initiated S $\gamma 1$ and S ϵ DSBs in anti-CD40/IL4-activated ATM-, H2AX-, and 53BP1-deficient B cells, as well as in B cells deficient for Rif-1, a 53BP1-associated factor that mediates resection blocking^{26,27}. ATM-, H2AX-, and Rif1-deficient B cells had reduced S $\gamma 1$ and S ϵ junctions compared to wild type; 53BP1-deficient B cells had a greater reduction, with most localizing to the break-site region (Fig. 3d, e and Extended Data Fig. 6b-d). Most break-site junctions were resections, which were longest (up to about 6 kb) for 53BP1 deficiency (Extended Data Fig. 6e, f; see extended discussion in Supplementary Information for Extended Data Fig. 6f). Compared to wild type, bait 5'S μ junctions to S $\gamma 1$ and S ϵ DSBs in different DSBR-deficient backgrounds had varying decreases in orientation specificity, with H2AX deficiency having the smallest and 53BP1 deficiency the largest (Figs 3d, e and 4a; Extended Data Fig. 6c, d and Extended Data Table 1a, b). Indeed, residual junctions of 5'S μ to S $\gamma 1$ and S ϵ locales in 53BP1-deficient B cells showed relatively normalized inversion:deletion ratios (Fig. 4a), a finding confirmed by DC-PCR (Extended Data Fig. 1b). Finally, 53BP1-deficiency did not impact joining orientation of 5'S μ and 3'S $\gamma 1$ I-SceI-generated broken ends in AID-deficient *Igh*^{I-96k} B cells (Extended Data Fig. 1g).

Owing to the potential difficulty in measuring relative resection of recurrent re-joins at or near the break-site, we focused on prey S-region broken end resections (Extended Data Figs 1 and 6; see extended discussion in Supplementary Information for Extended Data Figs 1d, e and 6f). Because S regions are long and AID-initiated DSB locations within them are diverse, we estimated relative resection

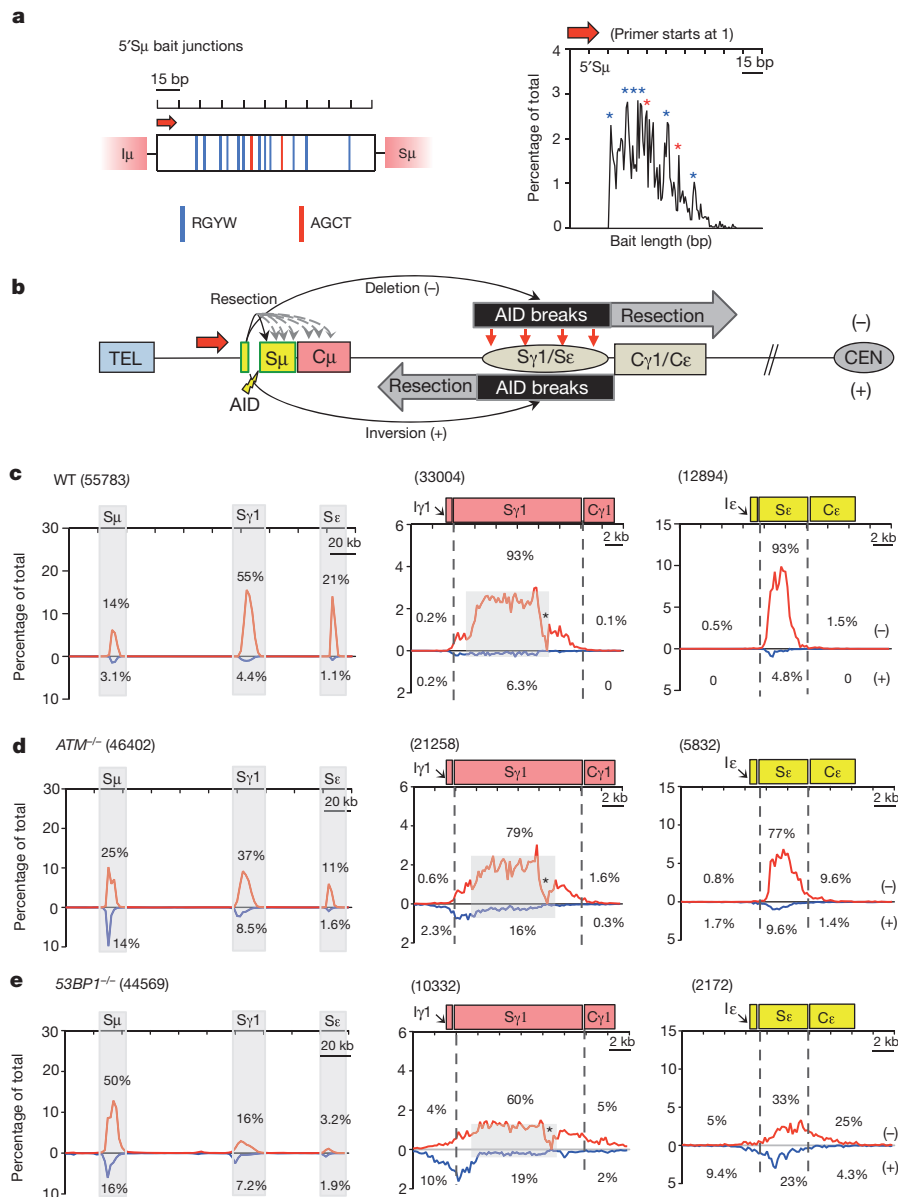


Figure 3 | Orientation-biased joining of AID-initiated endogenous S-region breaks. **a**, Left, 150-bp 5'Sμ sequence used as HTGTS bait. Red arrow denotes 5'Sμ primer. Red and blue vertical lines indicate AGCT or other AID-targeting motifs, respectively. Right, distribution and frequency of 5'Sμ break points in junctions to downstream S regions recovered from anti-CD40/IL4-stimulated wild-type B cells. Asterisks indicate positions of AGCT or other RGYW motifs. **b**, Junctional outcomes from 5'Sμ AID-initiated broken end joining to AID-initiated DSBs in Sγ1 and Sε including deletions (-) or inversions (+); long resections indicated by grey arrows. Break site 5'Sμ resections also are depicted. **c-e**, Linear distribution of pooled junctions along 200-kb C_H locus (left) or at Sγ1 and Sε (middle and right) recovered from anti-CD40/IL4-stimulated wild-type (n = 5) (c), *ATM*^{-/-} (n = 3) (d) or *53BP1*^{-/-} (n = 5) (e) B cells. Grey boxes indicate repetitive sequences with junctions mapping to multiple locations in Sγ1; asterisks indicate G-rich Sγ1 regions devoid of AID motifs and junctions.

by quantifying bait broken end to prey broken end junctions downstream of S-region positions where the incidence of wild-type junctions decreases to background (Fig. 3b-e). Based on this 'long' S-region resection assay, *ATM*- and *H2AX*-deficient cells had modest resection increases, *Rif1*-deficient cells slightly greater increases, and *53BP1*-deficient B cells far greater increases that were also apparent as a 'flattening' of Sγ1 and Sε junction profiles relative to other backgrounds (Figs 3c-e and 4b; Extended Data Fig. 6c, d and Extended Data Table 1c, d). HTGTS assays of rSμ bait broken end junctions to Sγ1 and Sε (Extended Data Fig. 7) and *I-SceI*-generated 3' ΔSγ1^{2X1} broken end bait junctions to Sμ and Sε (Extended Data Fig. 8) gave similar results. In *H2AX*- or *Rif1*-deficient B cells, a large fraction of 5'Sμ junctions were within S regions, with the main difference from wild type being a subset of junctions extending beyond S regions, probably reflecting extensive resection of broken ends not rapidly fused (Fig. 4b and Extended Data Fig. 6c, d). Treatment of *53BP1*-deficient activated B cells with *ATM* kinase inhibitor substantially diminished very long S-region resections, but did not restore orientation-dependent joining (Fig. 4a, b; Extended Data Table 1 and Extended Data Fig. 9a-f). This finding may reflect shorter resections in inhibitor-treated *53BP1*-deficient versus *ATM*-deficient B cells that are not revealed

by our long resection assay. Another possibility would involve a putative specialized role for *53BP1* in stabilizing synapsed S regions²³.

We demonstrate that CSR is mechanistically programmed to occur in a productive deletional orientation. Based on our findings, we propose a working model for orientation-specific CSR, in which a key component is the organization of S regions within topologically-associated domains (TADs) that promote their frequent S-region synapsis^{2,12,13} via Langevin motion^{2,13,28} (Fig. 4c). Within such TADs, we implicate additional *Igh*-specific organizational features, not yet fully elucidated, in playing a fundamental role in mediating synapsis in an orientation that promotes deletional joining (Fig. 4c). We find that functions of such organizational features are complemented by S regions, potentially associated with their ability to promote AID-initiated DSBs, multiple frequent DSBs, or both. Our studies also implicate DSB factors in enforcing this mechanism (Fig. 4d). The broader DSB factor probably contributes by tethering un-synapsed S-region DSBs for efficient re-joining, keeping them from separating into chromosomal breaks that could frequently translocate with orientation independence to S-region broken ends within the TAD^{2,20}; this function would also allow subsequent AID-initiated breakage and joining to a synapsed S region (Fig. 4c). DSB factors

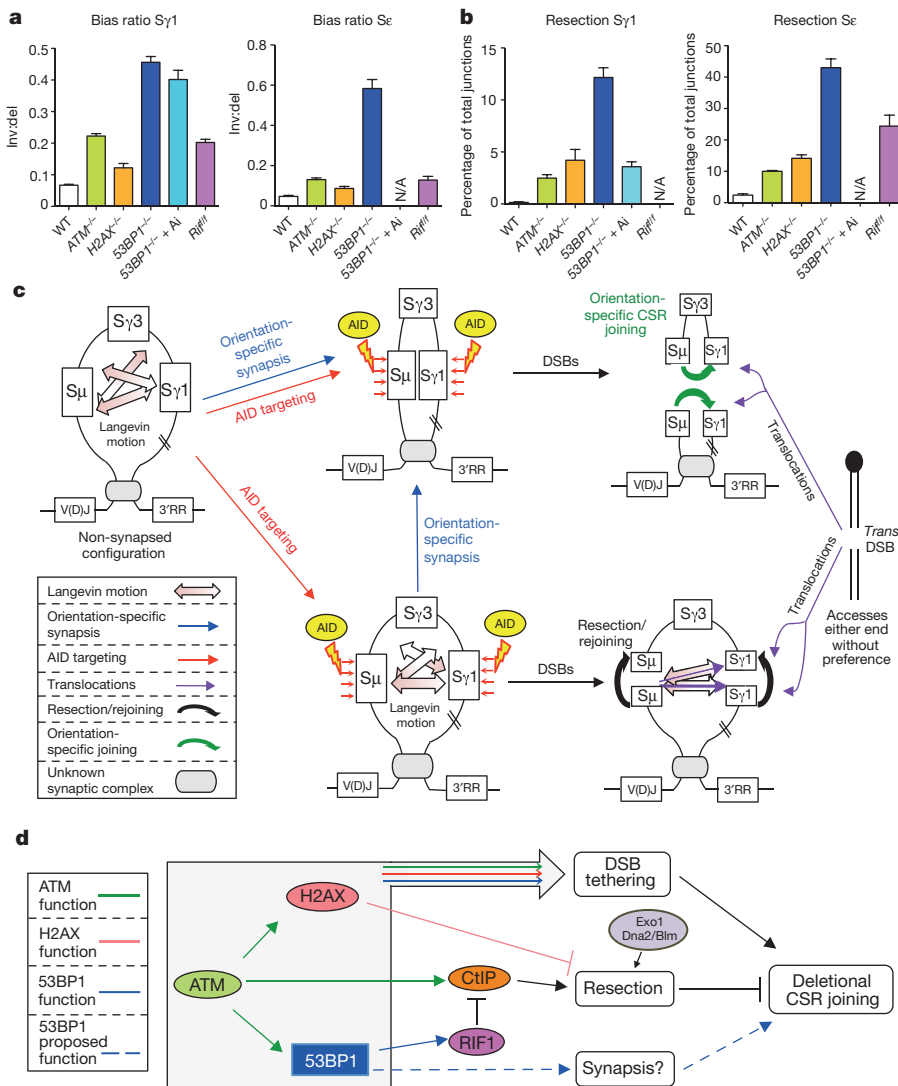


Figure 4 | Mechanistic roles of *IgH* organization and DSBR factors in deletional CSR. **a**, Ratios of inversional to deletional 5' S μ joins to S γ 1 (left) or S ϵ (right) in wild-type and mutant cells. Average \pm s.d. calculated from at least three separate experiments. **b**, S γ 1 and S ϵ resection junctions mapping to deletional (–) resection regions where the incidence of wild-type junctions decreases to background plotted as the percentage of total junctions in the deletional orientation from cells in **a**. For panels **a** and **b**, more than 1,000 unique junctions (up to tens of thousands in some cases) were analysed. Statistical significance or insignificance of key comparative results calculated by unpaired two-tailed *t*-test from at least three biological repeats (Extended Data Table 1). Ai: ATM inhibitor. **c**, **d**, Working model for orientation-biased joining (**c**) and functions of DSBR proteins (**d**) in maintaining directional CSR (details in text).

also prevent long end-resections that could cause S-region broken ends to linger in resection complexes, preventing synthesis with other S-region broken ends and/or diminishing ability to be joined by classical non-homologous end joining (Fig. 4d). Different DSBR factors have differential impact in tethering versus resection inhibition and, thus, may impact orientation dependence via different routes. For example, ATM deficiency inhibits resection by impairing CtIP activation²⁹, but promotes resection via other nucleases by impairing inhibitory activities of H2AX, 53BP1 and, indirectly, Rif1^{26,27} (Fig. 4d). 53BP1-deficiency is unique in that it both impairs tethering for rejoining and activates resection of un-joined ends by failure to activate Rif1, leading to extreme resections and the greatest impairment of CSR and orientation-dependent joining (Fig. 4d). As common and unique impacts of 53BP1 deficiency markedly affect both donor and acceptor S regions, they would be multiplicative and, thereby, explain the profound impact of 53BP1-deficiency on CSR.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 7 April; accepted 21 July 2015.

Published online 26 August 2015.

- Schatz, D. G. & Swanson, P. C. V(D)J recombination: mechanisms of initiation. *Annu. Rev. Genet.* **45**, 167–202 (2011).
- Alt, F. W., Zhang, Y., Meng, F. L., Guo, C. & Schwer, B. Mechanisms of programmed DNA lesions and genomic instability in the immune system. *Cell* **152**, 417–429 (2013).

- Di Noia, J. M. & Neuberger, M. S. Molecular mechanisms of antibody somatic hypermutation. *Annu. Rev. Biochem.* **76**, 1–22 (2007).
- Chiarle, R. *et al.* Genome-wide translocation sequencing reveals mechanisms of chromosome breaks and rearrangements in B cells. *Cell* **147**, 107–119 (2011).
- Frock, R. L. *et al.* Genome-wide detection of DNA double-stranded breaks induced by engineered nucleases. *Nature Biotechnol.* **33**, 179–186 (2015).
- Yancopoulos, G. D. *et al.* Secondary genomic rearrangement events in pre-B cells: VHDJH replacement by a LINE-1 sequence and directed class switching. *EMBO J.* **5**, 3259–3266 (1986).
- Jäck, H. M. *et al.* Looping out and deletion mechanism for the immunoglobulin heavy-chain class switch. *Proc. Natl Acad. Sci. USA* **85**, 1581–1585 (1988).
- Vaandrager, J. W. *et al.* DNA fiber fluorescence *in situ* hybridization analysis of immunoglobulin class switching in B-cell neoplasia: aberrant CH gene rearrangements in follicle center-cell lymphoma. *Blood* **92**, 2871–2878 (1998).
- Lenz, G. *et al.* Aberrant immunoglobulin class switch recombination and switch translocations in activated B cell-like diffuse large B cell lymphoma. *J. Exp. Med.* **204**, 633–643 (2007).
- Harriman, W., Völk, H., Defranoux, N. & Wabl, M. Immunoglobulin class switch recombination. *Annu. Rev. Immunol.* **11**, 361–384 (1993).
- Bothmer, A. *et al.* 53BP1 regulates DNA resection and the choice between classical and alternative end joining during class switch recombination. *J. Exp. Med.* **207**, 855–865 (2010).
- Zarrin, A. A. *et al.* Antibody class switching mediated by yeast endonuclease-generated DNA breaks. *Science* **315**, 377–381 (2007).
- Gostissa, M. *et al.* IgH class switching exploits a general property of two DNA breaks to be joined in cis over long chromosomal distances. *Proc. Natl Acad. Sci. USA* **111**, 2644–2649 (2014).
- Tian, M. & Alt, F. W. Transcription-induced cleavage of immunoglobulin switch regions by nucleotide excision repair nucleases *in vitro*. *J. Biol. Chem.* **275**, 24163–24172 (2000).
- Yu, K., Chedin, F., Hsieh, C. L., Wilson, T. E. & Lieber, M. R. R-loops at immunoglobulin class switch regions in the chromosomes of stimulated B cells. *Nature Immunol.* **4**, 442–451 (2003).

16. Reynaud, S. *et al.* Interallelic class switch recombination contributes significantly to class switching in mouse B cells. *J. Immunol.* **174**, 6176–6183 (2005).
17. Khamlichi, A. A. *et al.* Immunoglobulin class-switch recombination in mice devoid of any $\text{S}\mu$ tandem repeat. *Blood* **103**, 3828–3836 (2004).
18. Zhang, T. *et al.* Downstream class switching leads to IgE antibody production by B lymphocytes lacking IgM switch regions. *Proc. Natl Acad. Sci. USA* **107**, 3040–3045 (2010).
19. Nussenzweig, A. & Nussenzweig, M. C. Origin of chromosomal translocations in lymphoid cancer. *Cell* **141**, 27–38 (2010).
20. Franco, S. *et al.* H2AX prevents DNA breaks from progressing to chromosome breaks and translocations. *Mol. Cell* **21**, 201–214 (2006).
21. Bredemeyer, A. L. *et al.* ATM stabilizes DNA double-strand-break complexes during V(D)J recombination. *Nature* **442**, 466–470 (2006).
22. Daniel, J. A. & Nussenzweig, A. The AID-induced DNA damage response in chromatin. *Mol. Cell* **50**, 309–321 (2013).
23. Reina-San-Martin, B., Chen, J., Nussenzweig, A. & Nussenzweig, M. C. Enhanced intra-switch region recombination during immunoglobulin class switch recombination in 53BP1^{-/-} B cells. *Eur. J. Immunol.* **37**, 235–239 (2007).
24. Bassing, C. H. & Alt, F. W. The cellular response to general and programmed DNA double-strand breaks. *DNA Repair (Amst.)* **3**, 781–796 (2004).
25. Yamane, A. *et al.* RPA accumulation during class switch recombination represents 5′–3′ DNA-end resection during the S-G2/M phase of the cell cycle. *Cell Rep.* **3**, 138–147 (2013).
26. Panier, S. & Boulton, S. J. Double-strand break repair: 53BP1 comes into focus. *Nature Rev. Mol. Cell Biol.* **15**, 7–18 (2014).
27. Zimmermann, M. & de Lange, T. 53BP1: pro choice in DNA repair. *Trends Cell Biol.* **24**, 108–117 (2014).
28. Lucas, J. S., Zhang, Y., Dudko, O. K. & Murre, C. 3D trajectories adopted by coding and regulatory DNA elements: first-passage times for genomic interactions. *Cell* **158**, 339–352 (2014).
29. Helmink, B. A. *et al.* H2AX prevents CtIP-mediated DNA end resection and aberrant repair in G1-phase lymphocytes. *Nature* **469**, 245–249 (2011).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank K. Yu for providing the CH12F3-RMCE (1F7) cell line and exchange cassette plasmid. This work was supported by National Institute of Health grants AI077595 to F.W.A., CA133781 to J.M., AI112602 to D.F.R., and AI037526 and AI072529 to M.C.N. S.V. was supported by NIH training grant T32HL066987. F.W.A. and M.C.N. are investigators of the Howard Hughes Medical Institute. J.H. is supported by a Robertson Foundation/Cancer Research Institute Irvington Fellowship. F.M. is a Lymphoma Research Foundation postdoctoral fellow and was a Cancer Research Institute postdoctoral fellow.

Author Contributions J.D., R.P., T.Z., J.P.M. and F.W.A. designed the study; J.D., R.P., T.Z., J.H. and S.V. performed experiments; Y.H. and R.M. designed bioinformatics pipelines; R.M., J.D., R.P. and Z.D. performed computational analyses of sequencing data; J.D., R.P. and F.W.A. wrote the paper. Other authors provided reagents, designed or performed certain experiments, and helped revise the paper.

Author Information HTGTS sequencing data has been deposited in the GEO database under the accession number GSE71005. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to F.W.A. (alt@enders.tch.harvard.edu) or to J.P.M. (manis@enders.tch.harvard.edu).

METHODS

No statistical methods were used to predetermine sample size. The experiments were not randomized and the investigators were not blinded to allocation during experiments and outcome assessment.

Mice. *Igh*^{L-96k} *AID*^{-/-} (ref. 11), *ΔSμ*^{2×1/ΔSγ}^{12×1} chimera¹², *ΔSγ*^{12×1-Scel} (ref. 12), *Sμ*^{ΔH/Δ} (ref. 17), *c-myc*^{25×1-Scel} (ref. 4), *ATM*^{-/-} (ref. 30), *H2AX*^{-/-} (ref. 31), *53BP1*^{-/-} (ref. 32), and *Rifl*^{fl/fl} *CD19*^{Cre} (ref. 33) lines have been reported previously. Mouse work was performed under protocols approved by the Boston Children's Hospital and the Rockefeller University Institutional Animal Care and Use Committees.

Plasmids and oligonucleotides. Oligonucleotides for gRNAs for CRISPR/Cas9-mediated targeting of various *Igh* regions were cloned into pX330 vector (Addgene plasmid ID 42230) as described³⁴. The target sequences of Cas9 constructs are listed in the DNA oligonucleotides table in the Supplementary Information. Exchange vector (pLH28) with heterologous *loxP* sites was obtained from K. Yu. A 200-bp GFP-derived sequence was amplified and ligated to an *I-SceI* recognition sequence and subsequently introduced into the pLH28 vector to make the pLH-1× *I-SceI* exchange vector. To obtain the *I-SceI* expression plasmid for transducing CH12 cell lines, *I-SceI*-IRES-GFP fragment was shuttled from a retroviral construct (pMX-I-SceI-IRES-GFP) into pCDNA3.0 (Invitrogen) vector.

B-cell culture, transduction and FACS analysis. Mature splenic B cells isolated using a CD43-negative selection kit (MACS) were cultured in lymphocyte medium R15 (RPMI1640, 15% FBS, l-glutamate, 1× penicillin and streptomycin). B-cell stimulation was performed with anti-CD40 (1 μg ml⁻¹, eBioscience) plus IL4 (20 ng ml⁻¹, PeproTech) or LPS (25 ng ml⁻¹, Sigma) plus anti-IgD-dextran (3 ng ml⁻¹, a gift from R. Casellas) for 96 h. Infection with *I-SceI* expression or control retrovirus was carried out at day 1 post-stimulation by the standard spinning method with the presence of 4 μg ml⁻¹ polybrene as previously described¹³. Efficiency of retrovirus infection and switching levels were evaluated by flow cytometry as previously described¹³. Where indicated, ATM inhibitor KU-55933 (Tocris) was added to stimulated cells at day 1 post-stimulation to a final concentration of 10 μM and was maintained during the course of the experiment until collection of the cells for FACS and HTGTS libraries.

Cell lines and nucleofection. CH12F3 cell line stimulation to IgA was performed as described³⁵. CH12F3 cells with recombinase-mediated cassette exchange (RMCE) in place of the endogenous *Sα* region, referred to as 1F7 cells³⁵ were maintained at 37 °C, 5% CO₂ and cultured in RPMI media with 10% FCS, 0.5% penicillin/streptomycin, 50 μM β-mercaptoethanol. Exchange vector with heterologous *loxP* sites containing 1× *I-SceI* site embedded in 200 bp of GFP-derived sequence was cloned. RMCE was performed as previously described³⁵. Exchanged *ΔSα*^{1×1} clones were verified by PCR, Sanger sequencing and Southern blotting. *ΔSα*^{1×1} cells were then stimulated with anti-CD40, IL4 and TGF-β for 15 h followed by nucleofection with pCDNA-I-SceI-IRES-GFP expression vector using 4D-nucleofector X (Lonza, solution SF, protocol CA-137) and re-plated in stimulation-conditioned media. On day 3 post-stimulation cells were collected and gDNA was isolated for HTGTS library preparation. Cells were not tested for mycoplasma contamination.

To obtain CH12F3 (productive allele *Sμ*(*INV*), non-productive allele *ΔSμ*-*Sα*) cells, wild-type CH12F3 cells were first nucleofected using the 4D-nucleofector X (Lonza, solution SF protocol CA-137) with the gRNA vectors to excise the sequences between J_H4 intron and ~130 bp downstream of Cα polyadenylation on the non-coding allele that has already switched to *Sα*. Single-cell subclones were seeded into 96-well plates 12 h post-nucleofection, and the resulting clones were screened by PCR and Southern blot. One confirmed positive clone was further modified by gRNA vectors targeted at the 5' *Sμ*₁ and 3' *Sμ* regions to invert the *Sμ* (~4 kb) sequence. Initial screening for positive clones was performed by PCR, followed by Southern blotting and Sanger sequencing for the inversion junction. The resultant cells were stimulated with anti-CD40, IL4 and TGF-β, IgA CSR was measured by FACS on days 2 and 3 post-stimulation. *ΔSα*^{1×1} *Sμ*(*INV*) cells were obtained by targeting the aforementioned 1× *I-SceI* RMCE-positive cells with gRNA targeting 5' *Sμ*₂ and 3' *Sμ* for inverting the *Sμ* sequence same as above. The resultant positive clones were verified by PCR, Southern blotting and Sanger sequencing for the inversion junction. To make r*Sμ*-CH12F3 cells, the aforementioned CH12F3 (non-productive allele *ΔSμ*-*Sα*) cells were used to further truncate *Sμ* sequences on the coding allele with gRNA targeting 5' *Sμ*₂ and 3' *Sμ*. Single-cell deletion subclones were screened and confirmed by PCR and Southern blot. The resultant r*Sμ*-CH12F3 cells were stimulated with anti-CD40, IL4 and TGF-β and harvested on days 2 and 3 for gDNA isolation for HTGTS library preparation.

DC-PCR. The DC-PCR assay was performed as described previously³⁶. In brief, genomic DNA was isolated and subsequently purified by phenol chloroform extraction from day 4 anti-CD40/IL4 stimulated B cells. Five micrograms of genomic DNA was digested overnight with 20 U of EcoRI (Roche). Ligations were performed under diluted conditions to promote circularization. Digested DNA

was ligated overnight at 16 °C with a concentration of 1.8–9 ng μl⁻¹ in a total volume of 100 μl per reaction. Three to four ligation reactions were pooled, column purified, concentrated and serially diluted at a 1:5 ratio. PCR was then performed in 50 μl per reaction using 2.5 U Taq (Qiagen) with serially diluted DNA starting from ~50–150 ng. Primers were designed to amplify the *Sμ*-*Sγ*1 rearrangements that occur during CSR to IgG1 in direct chromosomal joining of *Sμ*-*Sγ*1 with excision of circular DNA or inversion of sequences between broken ends of *Sμ* and *Sγ*1. As a control for EcoRI digestion and circularization of input DNA, amplification of an EcoRI fragment of nicotinic acetylcholine receptor B subunit gene (*CHRNA1*) was performed, which, after EcoRI digestion and circularization, generates a 753-bp DC-PCR product. To quantify the amount of direct or inversion joins amplified by PCR, DC-PCR products of direct or inversion joins were cloned into the pCR2.1 Topo TA vector. Precise plasmid concentrations were determined and a standard curve was generated ranging from 4 to 10,000 copies per reaction. After running on 1% agarose gel, PCR fragments were transferred to nitrocellulose membrane and hybridized to a 3' *Sγ*1 probe according to standard Southern blotting procedures. Primers for direct joining PCR: forward, 5'-CAT GAGAGCTGGAGCTAGTATGAAGGTG-3'; reverse, 5'-ACTGACTGACTGAGTGTCTCTCAAC-3'. Primers for inversional joining PCR: forward, 5'-CAG TCACAGAGAACTGATCCAGGTGAG-3'; reverse, 5'-CCATAGCAGTTGG TCAATCCTTGTCTCC-3'. Primers for control *CHRNA1* DC-PCR³⁶: forward, 5'-GCGCCATCGATGGACTGCTGTGGGTTTCAACCCAG-3'; reverse, 5'-GGC CGGTCGACAGCGCGCACTGACACCCTAAG-3'. Oligonucleotide probe for the detection of both deletional and inversional CSR joining products: *Sγ*1-CCTGGGTAGTTACAGGTCAAGGCT.

High-throughput genome-wide translocation sequencing (HTGTS). HTGTS libraries were generated by emulsion-mediated PCR (EM-PCR) and linear-amplification-mediated PCR (LAM-PCR) methods as described in ref. 5. In brief, sonicated (Bioruptor, Diagenode) gDNA was subjected to LAM-PCR using 1 U Taq polymerase (Qiagen) per reaction with a single biotinylated primer for 50 cycles of 94 °C for 180 s; 94 °C for 30 s; 58 °C for 30 s; 72 °C for 90 s. One more unit of Taq polymerase was added to the reaction mixture to execute PCR for an additional 50 cycles. Biotinylated DNA fragments were captured with Dynabeads MyOne streptavidin C1 beads (Invitrogen) at room temperature for 1 h, followed by on-bead ligation at 25 °C for 2 h with bridge adapters in the presence of 15% PEG-8000 (Sigma) and 1 mM hexamine cobalt chloride (Sigma). After washing beads with B&W buffer as described by the manufacturer, ligated products were subjected to 15 cycles of on-bead PCR with Phusion polymerase (Fisher), locus-specific and adaptor primer followed by blocking digestion with appropriate restriction enzymes to remove uncut germline gDNA. A third round of tagging PCR to add Illumina Miseq-compatible adapters at 5' and 3' ends of the second-round PCR product was carried out for another 10 cycles with Phusion polymerase. PCR products were size-fractionated for DNA fragments between 300–1000 bp on a 1% agarose gel, column purified (Qiagen) before loading onto Illumina Miseq machine for sequencing.

Data analyses. Data analysis of MiSeq sequencing reads has been described in ref. 5. In brief, de-multiplexing for the MiSeq reads was performed using the fastq-multiplex tool from ea-utils (<https://code.google.com/p/ea-utils/>) and adaptor sequence trimming was performed using the SeqPrep utility (<https://github.com/jstjohn/SeqPrep>). Reads were mapped using Bowtie2 (<http://bowtie-bio.sourceforge.net/bowtie2/manual.shtml>) to either mm9 (for libraries generated with *Rifl* knockout cells and CH12F3-derived cells) or modified mm9 reference genome (for all other genotypes) containing the 176-kb *Igh* constant region of 129S genome, in which the region between chr12:114493849–114465808 of mm9 was replaced with DNA sequence ranging from 1416975 to 1593283 on the 129S *Igh* reference sequences AJ851868.3. In cases where necessary, for instance when aligning reads to the *Sμ*^{1×1} locus on the *Igh*^{L-96k} allele and other circumstances, we further modified the custom 129S₁-IgHC genome to insert the cassette sequences to accurately reflect the changes of genomic information before aligning MiSeq reads by Bowtie2. CH12F3 clone was derived from CH12.LX lymphoma cell line³⁷. CH12.LX cells were subcloned from the original CH12 lymphoma cell line³⁸, which originated from a C57BL/10 mouse substrain double congenic for H-2^a H-4^b (ref. 39). C57BL/10 and C57BL/6 are both substrains of C57BL and thus we use BL/6 (mm9) as reference genome when running our HTGTS data analyses pipeline on libraries made with CH12F3 cells. To reflect additional genome modifications (for example, *Sμ*(*INV*) shown in Fig. 2b), the mm9 genome sequence was modified accordingly.

A best-path searching algorithm (based on YAHA read aligner and break point detector⁴⁰) was used to select optimal sequence alignments from Bowtie2-reported top alignments with an alignment score above 50, which represents a perfect 25-nucleotide (nt) local alignment. To avoid detecting possible mis-priming events, we set a bait alignment threshold of at least ten perfectly aligned nucleotides extending from the end of cloning primer. Aligned reads were subsequently

filtered on following criteria: (1) reads must include both a bait alignment and a prey alignment; and (2) the bait alignment cannot extend more than 10 nt beyond the targeted site. For reads mapped to the repetitive low-mappability regions, multiple competing alignments with identical or similar scores exist and the coordinates for best alignment are randomly chosen among the competing ones. For junctions mapped to each individual repetitive S region, there are no competing alignments from outside of that region as shown by simulation (see details below), although the exact junction coordinate within the region could not be identified. We also applied filter to remove duplicates (referred to as 'de-dup' hereafter) wherein the coordinates of the end of the bait alignment were compared to the start of the prey alignment across all reads. A read is marked as a duplicate if it has bait and prey alignment coordinates within 2 nt of another read's bait and prey alignments. To plot all the S-region junctions, we took the ones filtered by a mappability filter but unequivocally mapped to S regions and removed the repeats through the de-dup program mentioned above, before combining with 'good' reads passing both the mappability and de-dup filters. A grey box over S regions (for example, S_μ and S_{γ1}) in the figures is used to denote the repetitive regions in these S sequences wherein the randomly assigned mappability-filtered reads were included. Additionally, we applied post-filtering stringencies to remove junctions mapped to simple sequence repeats, telomere repeats and reads with excessive microhomology >20 nt and insertions >30 nt before further analysis. In the end, the combined and cleaned junctions were then plotted genome-wide or onto desired S regions by using the PlotRegion tool (for details see section below).

Scripts and details of pipeline parameters are available upon request.

Pipeline simulation for S-region mappability. Results of the S-region mappability simulations are available in the Supplementary Information.

S-region junction plotting. As described above, junctions filtered by the mappability filter are retrieved and de-duped before combining with normal junctions. To plot junction coordinates onto individual S regions or the entire *Igh* constant region, combined junctions are binned using the PlotRegion tool into 100 bins (bin size varies depending on the length of target region that libraries are plotted to) on the basis of the junction coordinates and orientation of joining. The bincount file (histogram information for junction distribution in both joining orientation) generated by the PlotRegion tool is used to calculate the percentage of junctions in each bin in either + or - orientation of the total number of junctions mapped to the region of interest. The results were then plotted as linear graphs by the Prism software. Note that the scale on top of each graph indicates the size of region plotted and is fixed as 1/10 of the size of the plotted region, thus is always 10× bin size.

Calculation of joining orientation bias and acceptor S-region resection. For simplicity, joining from 5' S_μ to downstream S_{γ1} and S_ε breaks are used for the explanation of orientation bias and resection of acceptor S-region DSBs. Junctions mapped to S_{γ1} and S_ε can be divided into six regions (denoted by *a-f*) in either + or - orientation:

$$\frac{a|b|c(-)}{d|e|f(+)}$$

Junctions encompassing core S_{γ1}/S_ε are illustrated as *b* and *e* regions for - and + junctions respectively, *c* region (deletional joining, - orientation) or *d* region

(inversional joining, + orientation) represent joining of bait DSB broken ends to resected acceptor S_{γ1}/S_ε DSBs. Junctions falling into regions *a* or *f* represent joining to non-AID-generated *de novo* breaks of unknown source and are often very small in number, and thus were omitted from the calculation of both resection and orientation bias. Since in most genetic backgrounds other than 53BP1^{-/-} inversion joins are much rarer than deletions, the level of resection junctions into the *d* region fluctuates much more than resection junctions into the *c* region. We thus chose the *c* region for calculating resection in all genotypes as follows:

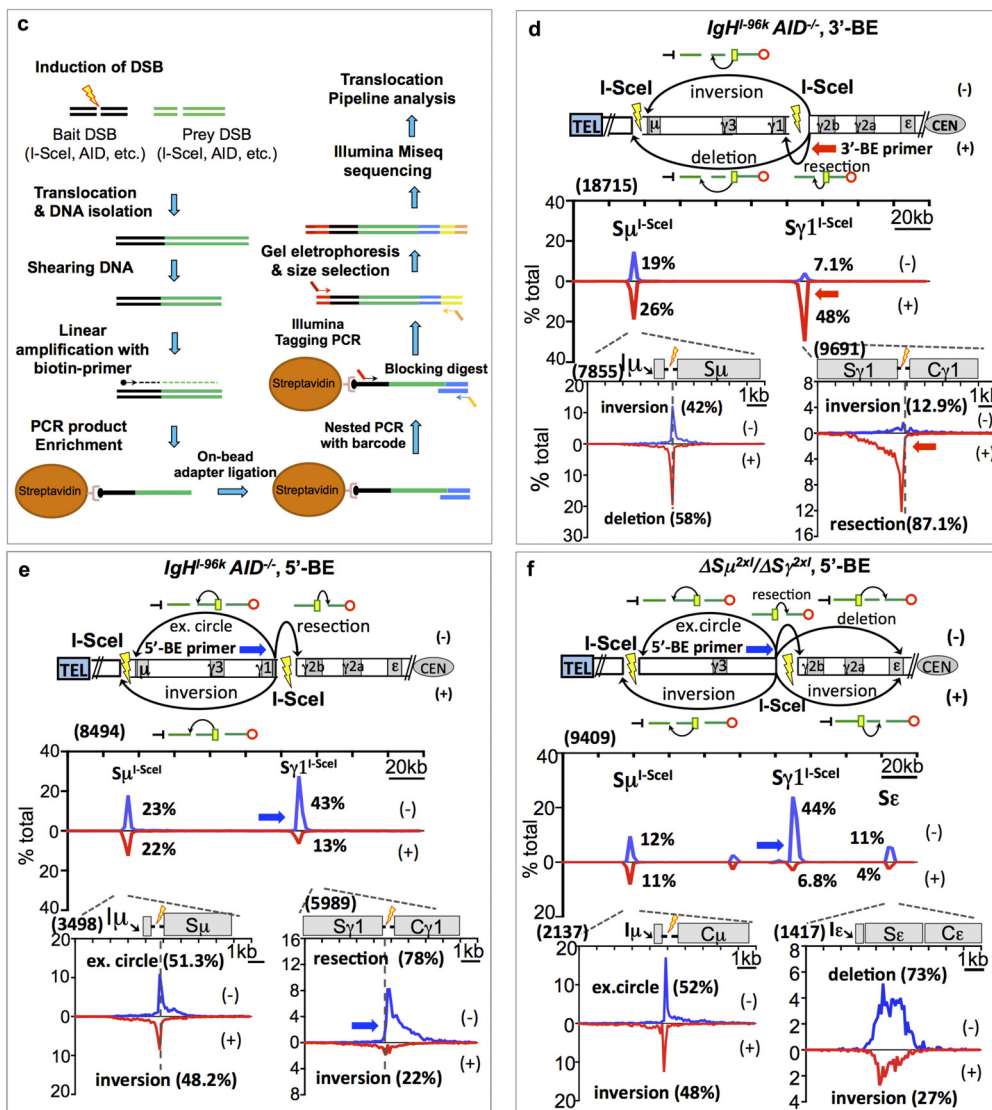
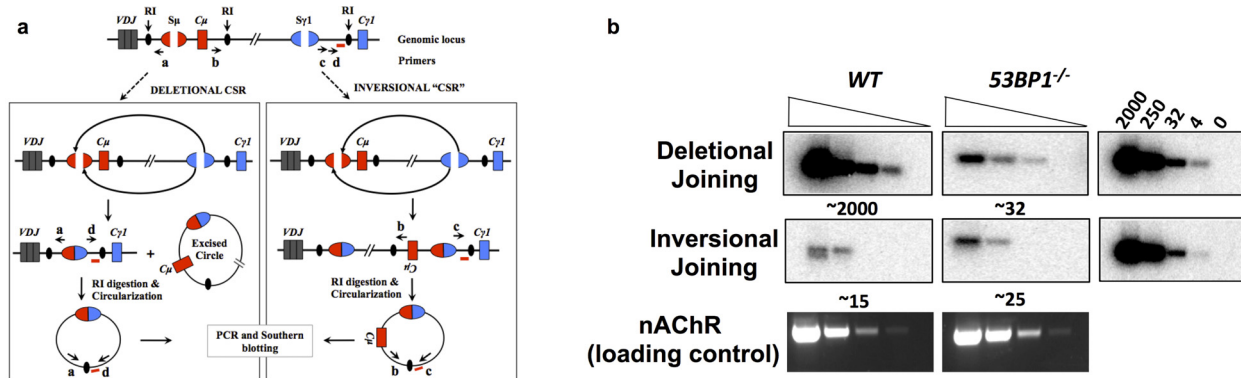
$$\text{resection rate} = \frac{c}{b+c} \times 100$$

The degree of orientation bias, for the purpose of positively correlating with the level of resection, is calculated as the ratio of inversional joins versus deletional joins as below:

$$\text{bias ratio} = \frac{d+e}{b+c} \times 100$$

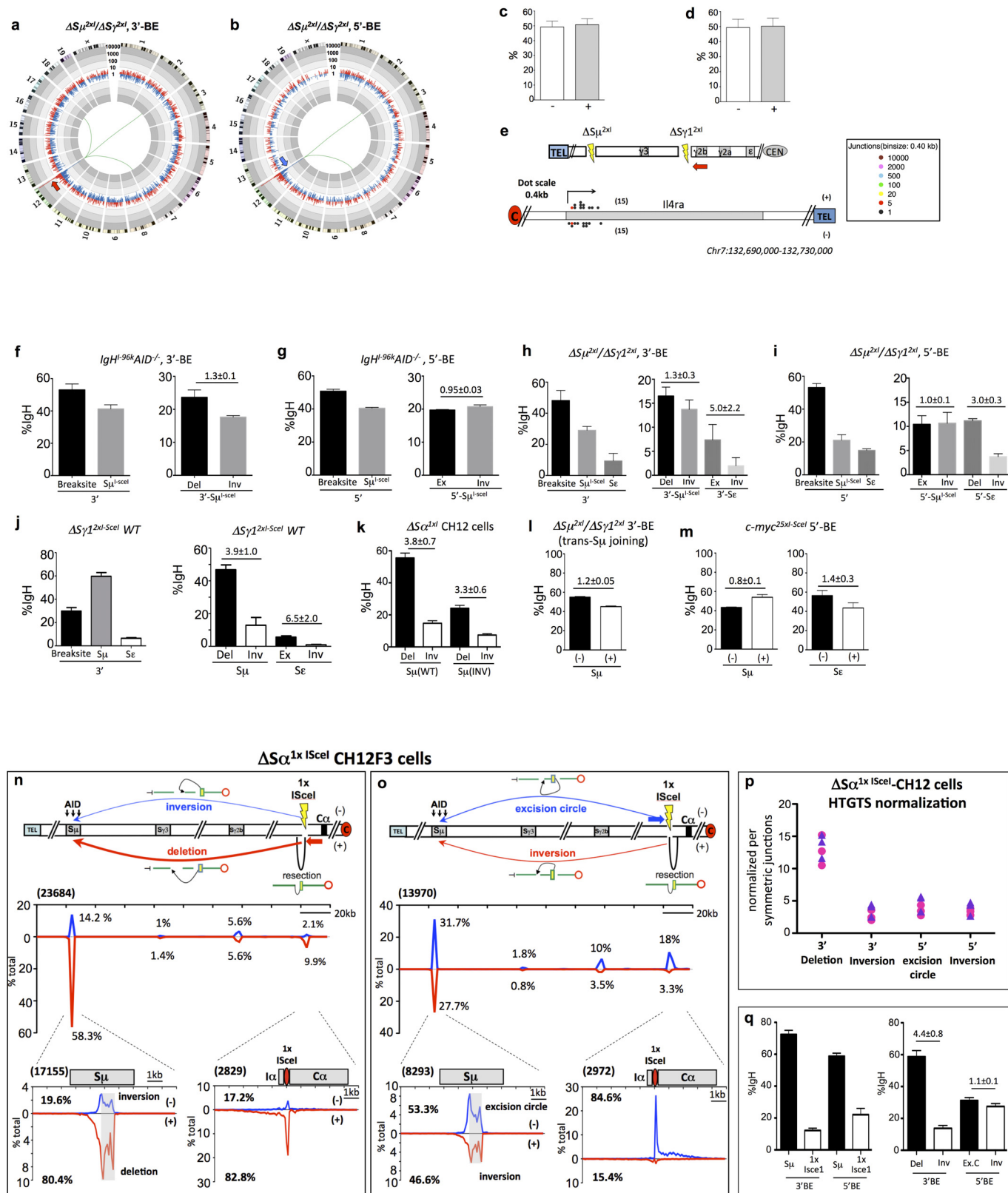
To make a bar graph for comparison of orientation bias degree and resection levels in the CSR junctions obtained from libraries with different genetic backgrounds, individual replicate HTGTS libraries were first size-normalized to the one with smallest junction number in the region of interest among the replicates; resection and bias ratio values from individual experiments were calculated separately and averages were used for statistical analysis with unpaired two-tailed *t*-tests. Experiments for each genotype were performed at least three times.

30. Borghesani, P. R. *et al.* Abnormal development of Purkinje cells and lymphocytes in *Atm* mutant mice. *Proc. Natl Acad. Sci. USA* **97**, 3336–3341 (2000).
31. Bassing, C. H. *et al.* Histone H2AX: a dosage-dependent suppressor of oncogenic translocations and tumors. *Cell* **114**, 359–370 (2003).
32. Morales, J. C. *et al.* Role for the BRCA1 C-terminal repeats (BRCT) protein 53BP1 in maintaining genomic stability. *J. Biol. Chem.* **278**, 14971–14977 (2003).
33. Di Virgilio, M. *et al.* Rif1 prevents resection of DNA breaks and promotes immunoglobulin class switching. *Science* **339**, 711–715 (2013).
34. Cong, L. *et al.* Multiplex genome engineering using CRISPR/Cas systems. *Science* **339**, 819–823 (2013).
35. Han, L., Masani, S. & Yu, K. Overlapping activation-induced cytidine deaminase hotspot motifs in Ig class-switch recombination. *Proc. Natl Acad. Sci. USA* **108**, 11584–11589 (2011).
36. Chu, C. C., Paul, W. E. & Max, E. E. Quantitation of immunoglobulin μ-γ1 heavy chain switch region recombination by a digestion-circularization polymerase chain reaction method. *Proc. Natl Acad. Sci. USA* **89**, 6978–6982 (1992).
37. Nakamura, M. *et al.* High frequency class switching of an IgM⁺ B lymphoma clone CH12F3 to IgA⁺ cells. *Int. Immunol.* **8**, 193–201 (1996).
38. Bishop, G. A. & Haughton, G. Induced differentiation of a transformed clone of Ly-1 B cells by clonal T cells and antigen. *Proc. Natl Acad. Sci. USA* **83**, 7410–7414 (1986).
39. Haughton, G., Arnold, L. W., Bishop, G. A. & Mercolino, T. J. The CH series of murine B cell lymphomas: neoplastic analogues of Ly-1⁺ normal B cells. *Immunol. Rev.* **93**, 35–52 (1986).
40. Faust, G. G. & Hall, I. M. YAHA: fast and flexible long-read alignment with optimal breakpoint detection. *Bioinformatics* **28**, 2417–2424 (2012).



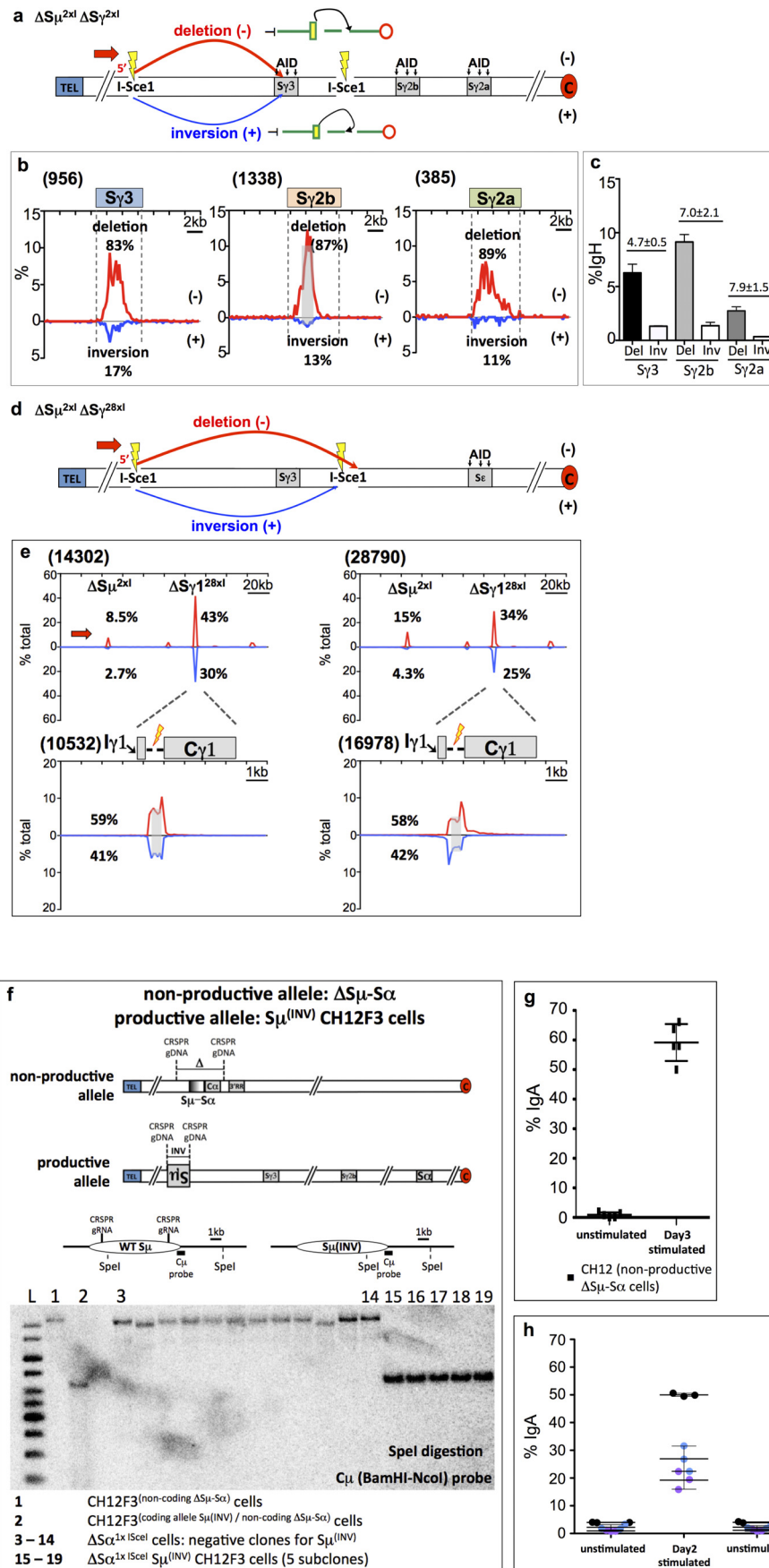
Extended Data Figure 1 | Deletional CSR in *in vitro* activated B cells by DC-PCR; I-SceI DSBs within the *Igh* constant region locus in activated B cells join with orientation-independence. a, Schematic representation of DC-PCR assay. **b**, DC-PCR results from anti-CD40/IL4-activated wild-type and 53BP1^{-/-} B cells. **c**, Schematic representation of the HTGTS method. **d, e**, HTGTS libraries analyses of anti-CD40/IL4-stimulated *Igh*^{I-96k} B cells with 3'-broken end (**d**, red arrow, *n* = 3) or 5'-broken end (**e**, blue arrow, *n* = 3)

primers. BE, broken end. **f**, HTGTS libraries with 5'-broken end primer (blue arrow, *n* = 3) from $\Delta S\mu^{2\times 1}/\Delta S\gamma^{2\times 1}$ B cells stimulated with anti-CD40/IL4. **g**, Bar graph depicting deletion:inversion and excision-circle:inversion ratios between two I-SceI sites and between I-SceI and S region in wild-type versus 53BP1^{-/-} backgrounds. For detailed legends and further discussion, refer to the Supplementary Information.



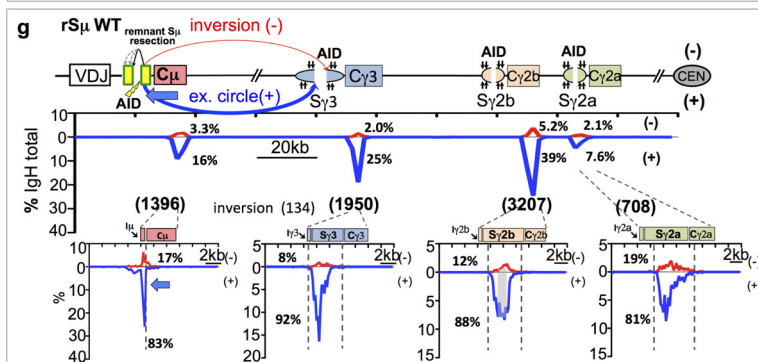
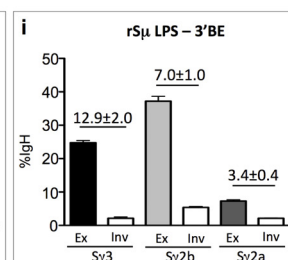
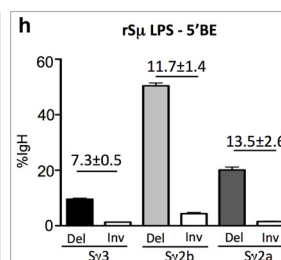
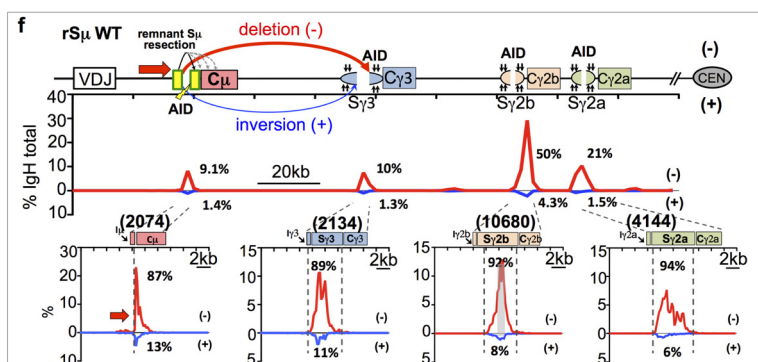
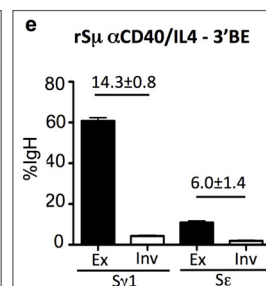
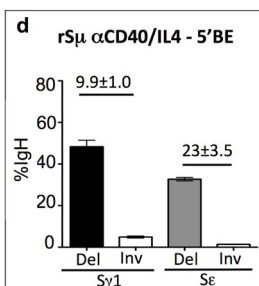
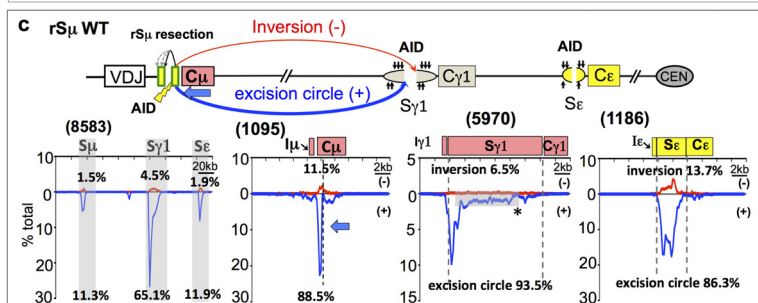
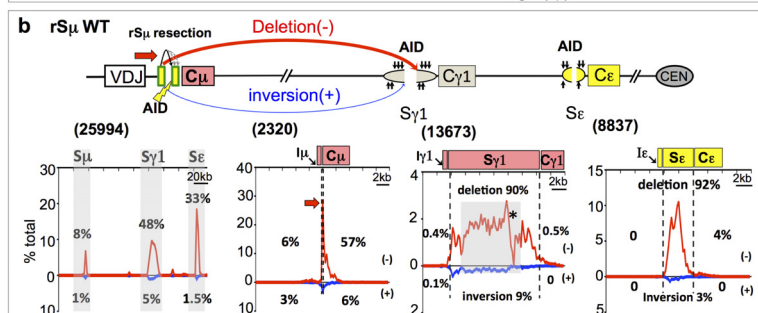
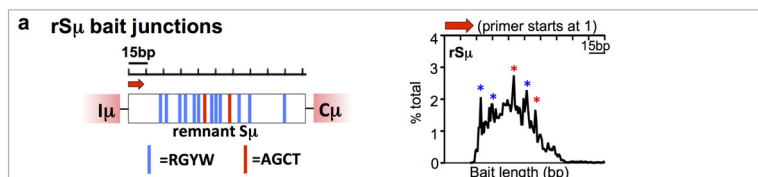
Extended Data Figure 2 | Genome-wide translocation junctions lack orientation bias; statistical analyses for experimental replicates orientation-biased joining between I-SceI break in place of S α and AID-initiated S μ breaks in CH12F3 cells. **a, b**, Circos plots for translocation junctions across the whole genome from 3'-broken end (**a**, $n = 4$) or 5'-broken end (**b**, $n = 3$) HTGTS with anti-CD40/IL4 stimulated $\Delta S\mu^{2\times I}/\Delta S\gamma I^{2\times I}$ B cells. **c, d**, Bar graphs depicting genome-wide percentage of junctions from pooled 3'- and 5'-broken end libraries plotted separately in – or + orientations. Error bars are s.d. **e**, Joining from $\Delta S\gamma I^{2\times I}$ 3'-broken end to AID off-target DSBs in *Il4ra* gene on chromosome 7. **f**, Bar graph showing the number of junctions (average \pm s.d.) recovered from *Igh*^{L-96k} *AID*^{-/-} 3'-broken end HTGTS libraries ($n = 3$) at the break site and the upstream S μ ^{1 \times I} prey break as a percentage of the total number of junctions mapped to the 200 kb *Igh* constant region. Right panel shows the percentage of junctions mapping at S μ ^{1 \times I} (average \pm s.d.) over the total *Igh* junctions that are mapped in the deletion (Del) or inversional (Inv) orientation. The numbers above the bar graph (average \pm s.d.) denote the ratio of deletional to inversional junctions. **g**, Percentage of junctions (average \pm s.d.) recovered from the *Igh*^{L-96k} *AID*^{-/-} 5'-broken end HTGTS libraries ($n = 3$). **h**, Percentage of junctions (average \pm s.d.) recovered from the $\Delta S\mu^{2\times I}/\Delta S\gamma I^{2\times I}$ 3'-broken end libraries

($n = 4$). **i**, Percentage of junctions (average \pm s.d.) recovered from the $\Delta S\mu^{2\times I}/\Delta S\gamma I^{2\times I}$ 5'-broken end libraries ($n = 3$). **j**, Percentage of junctions (average \pm s.d.) recovered from the wild-type $\Delta S\gamma I^{2\times I}$ 3'-broken end libraries ($n = 3$). **k**, Percentage of junctions (average \pm s.d.) recovered from the $\Delta S\alpha^{1\times I}$ CH12F3 3'-broken end libraries ($n = 3$) and $\Delta S\alpha^{1\times I}$ S μ (INV) CH12F3 cells 3'-broken end libraries ($n = 3$). **l**, Bar graphs depicting percentage of *trans* junctions mapping to S μ in – and + orientations from libraries of $\Delta S\mu^{2\times I}/\Delta S\gamma I^{2\times I}$ B cells ($n = 3$) cloning from $\Delta S\gamma I^{2\times I}$ 3'-broken ends. **m**, Bar graphs depicting percentage of *trans* junctions mapping to S μ in – and + orientations and to S ϵ in – and + orientations from libraries of *c-myc*^{25 \times I} 5'-broken ends ($n = 3$). **n, o**, HTGTS library analyses of $\Delta S\alpha^{1\times I}$ CH12F3 cells stimulated with anti-CD40, IL4 and TGF β and nucleofected with I-SceI expression plasmid. Cells were harvested on day 3 post-stimulation for 3'-broken ends (**n**, $n = 6$) and 5'-broken ends (**o**, $n = 6$) libraries. **p**, 3'- and 5'-broken end libraries are normalized with 'symmetric junctions' (see Supplementary Information). **q**, Bar graph showing percentage of junctions from $\Delta S\alpha^{1\times I}$ CH12F3 cells ($n = 6$) from 3'- and 5'-broken end primers. For detailed legends and further discussion refer to the Supplementary Information.



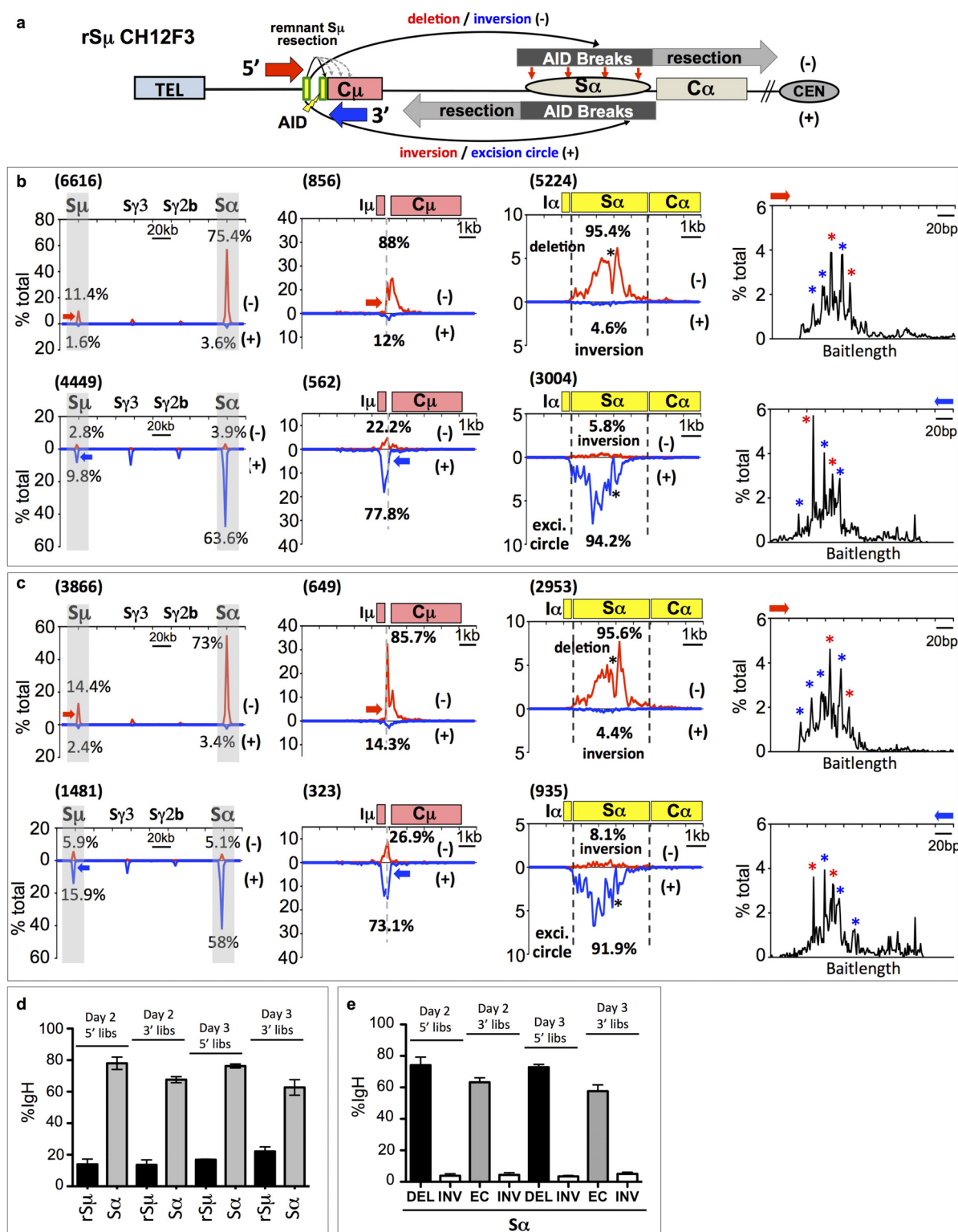
Extended Data Figure 3 | Joining between I-SceI break at S μ and AID-initiated S-region breaks in lipopolysaccharide (LPS)-activated $\Delta S\mu^{2\times 1}/\Delta S\gamma 1^{2\times 1}$ B cells and clustered I-SceI breaks in $\Delta S\mu^{2\times 1}/\Delta S\gamma 1^{28\times 1}$ B cells in place of S $\gamma 1$; inverted S μ in CH12F3 cells support robust IgA CSR. **a**, Diagram of *Igh* locus organization in $\Delta S\mu^{2\times 1}/\Delta S\gamma 1^{2\times 1}$ B cells highlighting AID-initiated breaks in S $\gamma 3$, S $\gamma 2b$ and S $\gamma 2a$ regions upon LPS stimulation and potential outcomes of CSR in the form of deletion (red, $-$) and inversional joining (blue, $+$). **b**, Plots showing enlarged distribution of pooled prey junctions in a 20-kb region flanking S $\gamma 3$ and S $\gamma 2b$ and S $\gamma 2a$ from HTGTS libraries of $\Delta S\mu^{2\times 1}/\Delta S\gamma 1^{2\times 1}$ B cells ($n = 3$) stimulated with LPS and anti-IgD-dextran and infected with I-SceI-expressing retrovirus. **c**, Bar graph from three independent $\Delta S\mu^{2\times 1}/\Delta S\gamma 1^{2\times 1}$ 5'-broken end libraries showing the percentage of junctions mapped at different S regions. **d**, Diagram of *Igh* locus

organization in $\Delta S\mu^{2\times 1}/\Delta S\gamma 1^{28\times 1}$ B cells highlighting joining outcomes of I-SceI-mediated bait DSBs at $\Delta S\mu^{2\times 1}$ to clustered I-SceI DSBs at $\Delta S\gamma 1^{28\times 1}$ in the form of deletion (red, $-$) and inversional joining (blue, $+$). **e**, Pooled prey junctions from independent $\Delta S\mu^{2\times 1}/\Delta S\gamma 1^{28\times 1}$ B cell libraries ($n = 2$, left panel, emulsion-mediated PCR; $n = 2$, right panel, linear-amplification-mediated HTGTS). **f**, Southern blot for S μ inversion on the productive allele of CH12F3 cells with non-productive allele deleted. **g**, IgA CSR on day 3 for CH12F3 cells stimulated with anti-CD40, IL4 and TGF β . **h**, IgA CSR on CH12F3 (productive allele *S μ (INV)*, non-productive allele *S μ (S α)*) cells stimulated with anti-CD40, IL4 and TGF β . Two independent clones of CH12F3 (*S μ (INV)*, non-productive *S μ (S α)*). For detailed legends and further discussion refer to the Supplementary Information.



Extended Data Figure 4 | Orientation-biased joining between AID-initiated rS μ and downstream AID-initiated S-region breaks in anti-CD40/IL4-activated and LPS-activated S μ -truncated B cells. **a**, 150-bp rS μ sequence used as HTGTS bait with red arrow indicating rS μ 5'-broken end HTGTS primer; red and blue vertical lines indicate canonical AGCT or other RGYW AID-targeting motifs, respectively. Distribution of rS μ break points in junctions to downstream S regions recovered from anti-CD40/IL4-stimulated rS μ B cells. **b**, HTGTS analyses of anti-CD40/IL4-activated rS μ B cells, 5' rS μ AID-initiated broken end (red primer, $n = 3$) junctions to AID-initiated DSBs in S γ 1 and S ϵ which includes deletion (– orientation, red) or inversions (+ orientation, blue). **c**, HTGTS analyses of anti-CD40/IL4-activated rS μ B cells, 3' rS μ AID-initiated broken end (blue primer, $n = 3$) junctions to AID-initiated DSBs in S γ 1 and S ϵ which includes excision circle (+ orientation, blue) or inversions (– orientation, red). **d**, Bar graph showing the percentage of junctions (average \pm s.d.) from anti-CD40/IL4-activated rS μ 5'-broken end

libraries mapped to S γ 1 and S ϵ . **e**, Bar graph showing the percentage of junctions (average \pm s.d.) from anti-CD40/IL4-activated rS μ 3'-broken end libraries mapped to S γ 1 and S ϵ . **f**, HTGTS analyses of LPS-activated rS μ B cells, 5' rS μ (red primer, $n = 3$) AID-initiated broken end junctions to AID-initiated DSBs in S γ 3, S γ 2b and S γ 2a which include deletional joining (– orientation, red) or inversions (+ orientation, blue). **g**, HTGTS analyses of LPS-activated rS μ B cells, 3' rS μ (blue primer, $n = 3$) AID-initiated broken end junctions to AID-initiated DSBs of above LPS-stimulated cells in S γ 3, S γ 2b and S γ 2a which include excision circle (+ orientation, blue) or inversions (– orientation, red). **h**, **i**, Percentage of junction distribution at S γ 3, S γ 2b and S γ 2a in both orientations from both 5'-broken end libraries (**h**) and 3'-broken end libraries (**i**) are shown as average \pm s.d. from three independent experiments. For detailed legends and further discussion refer to the Supplementary Information.



Extended Data Figure 5 | Orientation-biased joining between rSμ and AID-induced Sα DSBs in CSR-activated CH12F3 cells. **a**, Diagram outlining potential junction outcomes from 5' rSμ (red primer) or 3' rSμ (blue primer) AID-initiated broken end junctions to AID-initiated DSBs in Sα upon anti-CD40, IL4 and TGFβ stimulation of ΔSμ CH12F3 cells. **b**, **c**, Top panel shows HTGTS libraries analyses of day 2 (**b**) and day 3 (**c**) stimulated CH12F3 (non-productive allele ΔSμ-Sα, productive allele ΔSμ) cells cloning from 5'-broken end rSμ (red primer, $n = 3$), whereas lower panel shows HTGTS

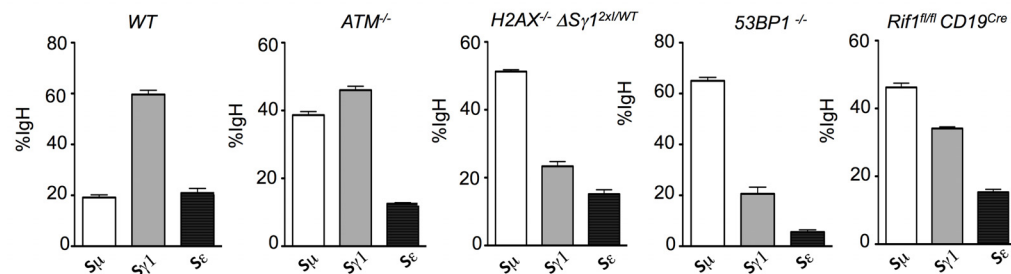
libraries cloning from 3'-broken end rSμ (blue primer, $n = 3$). **d**, Bar graph shows percentage of junctions (average ± s.d.) for 5'-broken end and 3'-broken end libraries indicated in **b** and **c**. **e**, Bar graph with percentage of junctions (average ± s.d.) from rSμ libraries mapped to prey Sα in the deletion (DEL) or inversion (INV) for 5'-broken end libraries and in excision circle (EC) or inversion orientation for 3'-broken end libraries. For detailed legends and further discussion refer to the Supplementary Information.

a

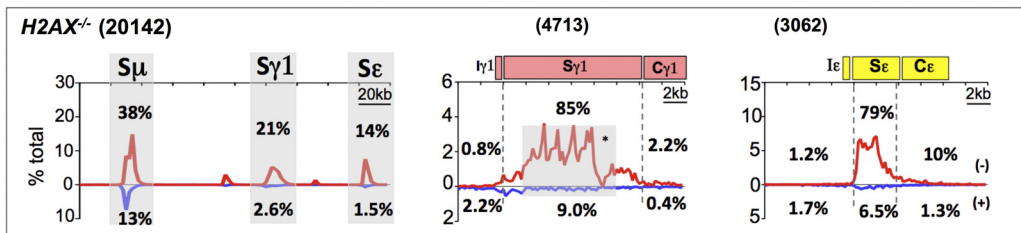
Day 4 CSR levels

	IgG1 (%)	IgE (%)
WT (n=5)	45% ± 2.8	14 ± 2
ATM ^{-/-} (n=3)	18.1 ± 0.6	N/A
H2AX ^{-/-} Sγ1 ^{1/2xISceI} (n=3)	8.3 ± 1.3	3.8 ± 0.3
53BP1 ^{-/-} (n=8)	3.2 ± 1.1	N/A
Rif1 ^{fl/fl} CD19 ^{Cre} (n=3)	6.6 ± 1.6	N/A

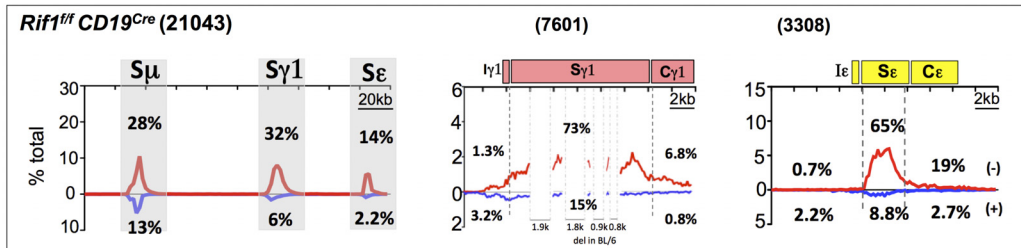
b



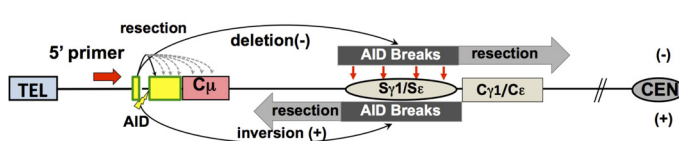
c



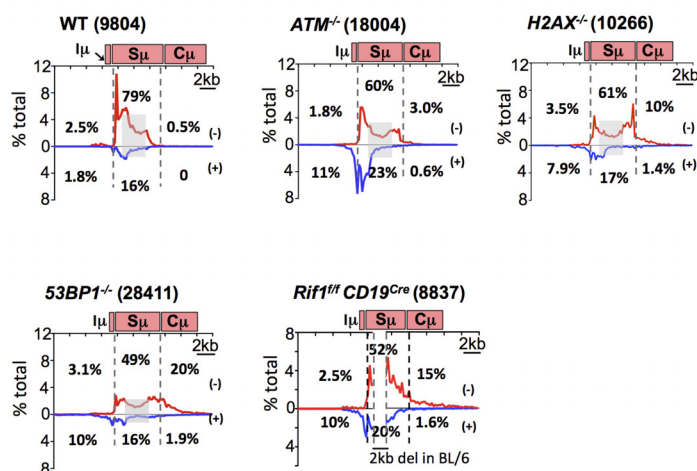
d



e

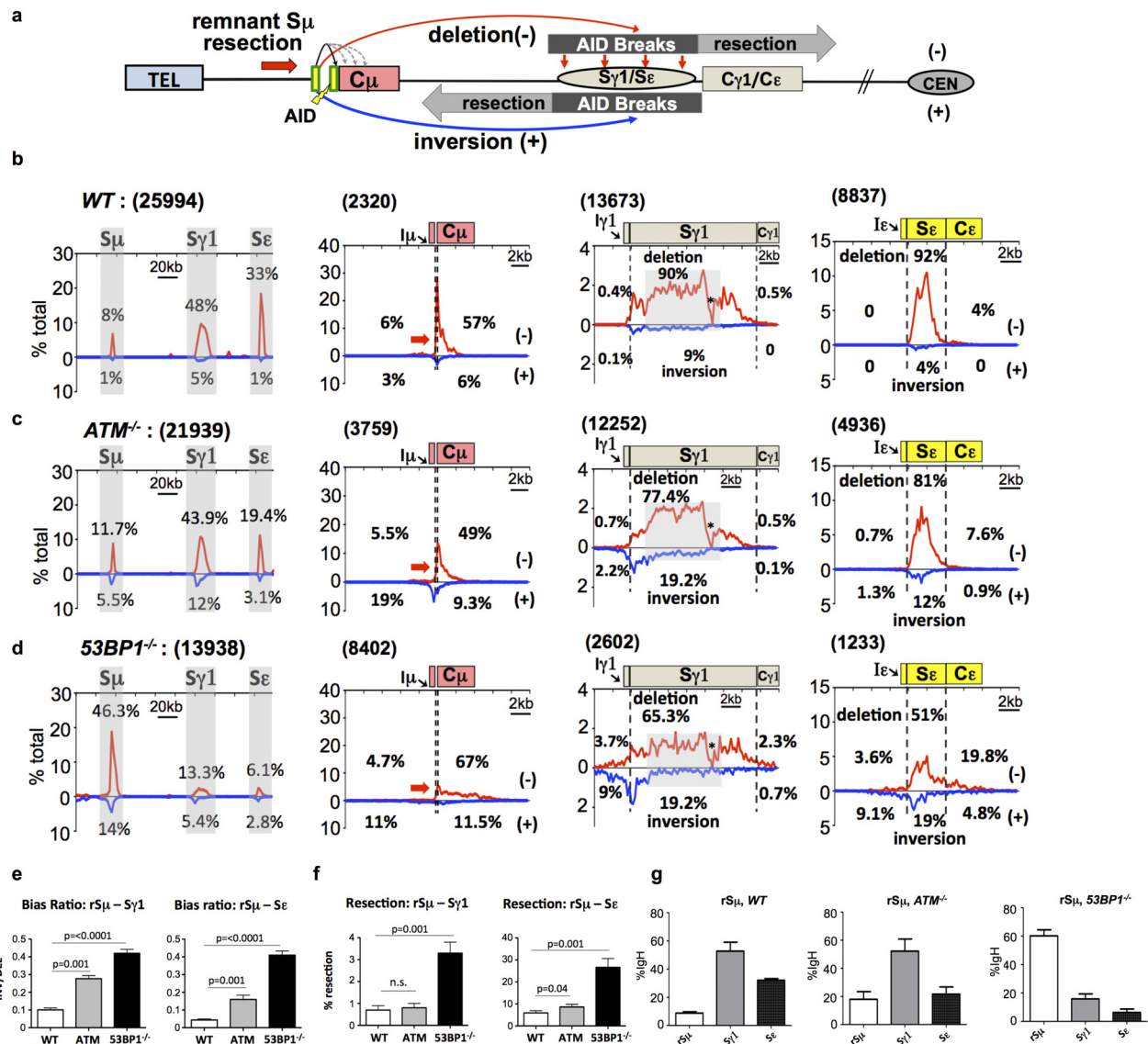


f



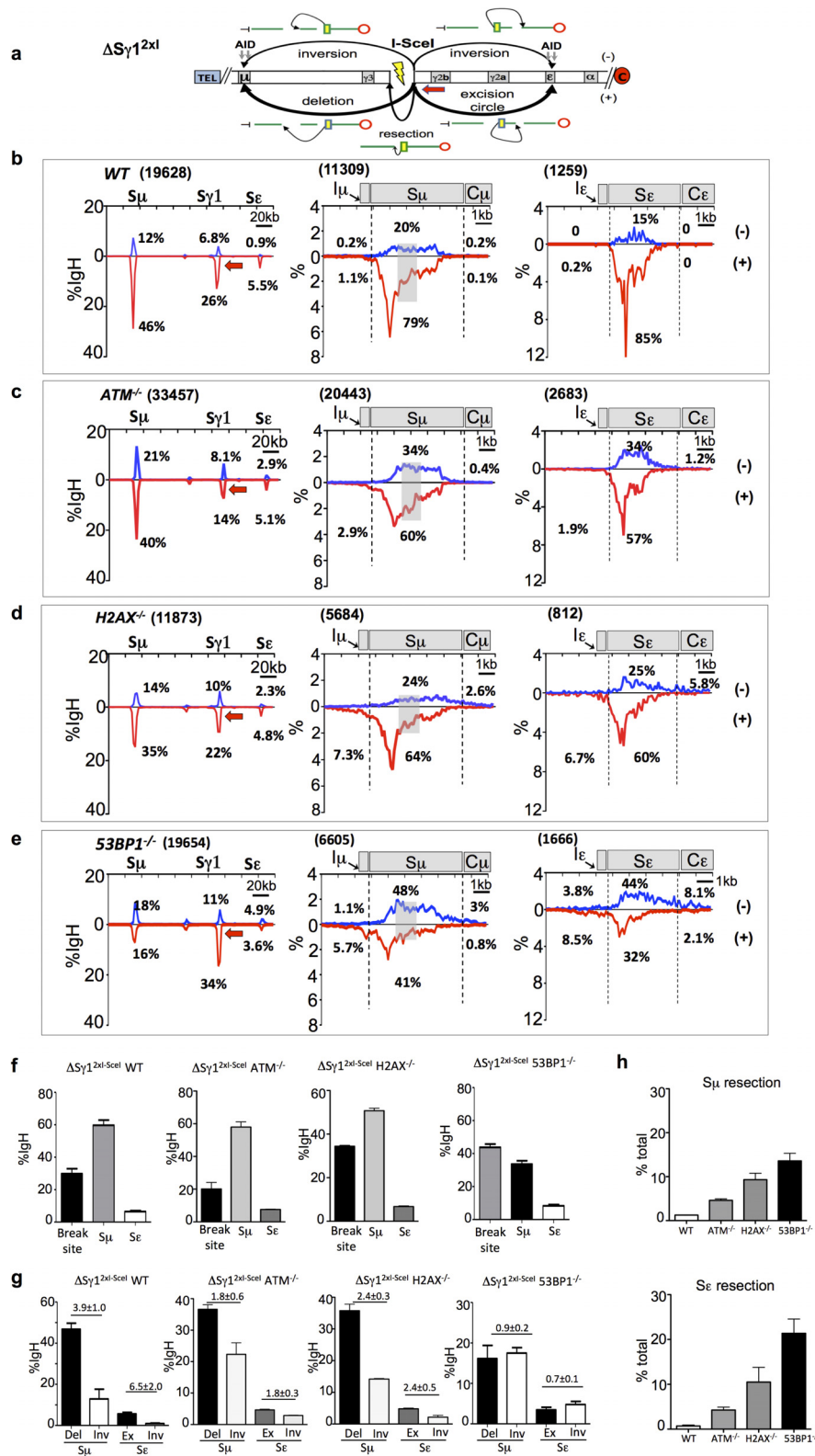
Extended Data Figure 6 | Level of junctions to downstream S regions in wild-type and DSBR-deficient 5'S μ HTGTS libraries correlate with CSR levels; 5'S μ break site undergoes variable degrees of resection from stimulated DSBR-deficient B cells. **a**, Table showing IgG1 and IgE CSR levels of splenic B cells from various genotypes (with number of replicates indicated) activated *in vitro* with anti-CD40 and IL4. FACS was performed on day 4 and values indicate average \pm s.d. WT, wild type. **b**, Left panel shows bar graph for percentage of junctions (average \pm s.d.) recovered from wild-type 5'S μ 5'-broken end libraries mapped to S μ , S γ 1 and S ϵ over the total number of junctions identified from the 200-kb *Igh* constant region. Remaining

panels show the similar results from different DSBR-deficient backgrounds using the same 5'-broken end primer. **c**, **d**, 5'S μ 5'-broken end HTGTS libraries analyses from *H2AX*^{-/-} and *RIFI* ^{β/β} *CD19*^{cre} B cells are shown respectively. **e**, Diagram of potential junction outcomes from 5'S μ AID-initiated 5'-broken end junctions to AID-initiated DSBs in S γ 1 and S ϵ . **f**, Data from HTGTS libraries mapped to the 20-kb region flanking 5'S μ break site from B cells stimulated with anti-CD40/IL4 in wild-type and DSBR-deficient backgrounds. For detailed legends and further discussion refer to the Supplementary Information.



Extended Data Figure 7 | Orientation-biased joining between rSμ and AID-induced Sγ and Sε DSBs in wild-type, ATM-deficient, and 53BP1-deficient B cells. **a**, Diagram of potential junction outcomes from 5' rSμ AID-initiated broken end junctions to AID-initiated DSBs in Sγ1 and Sε as described earlier. **b–d**, Linear plots of pooled junctions across the 200-kb *Igh* constant region (first panel), the 20-kb region flanking rSμ break site (second panel), the 20-kb region flanking downstream Sγ1 (third panel) and Sε (last panel) from three independent wild-type (**b**), ATM- (**c**), or 53BP1- (**d**) deficient 5'-broken end rSμ libraries. **e**, Bar graphs showing inversion:deletion (INV/DEL) bias ratios of HTGTS Sγ1 (left panel) and Sε (right panel) junctions in different genotypes, showing average \pm s.d. from

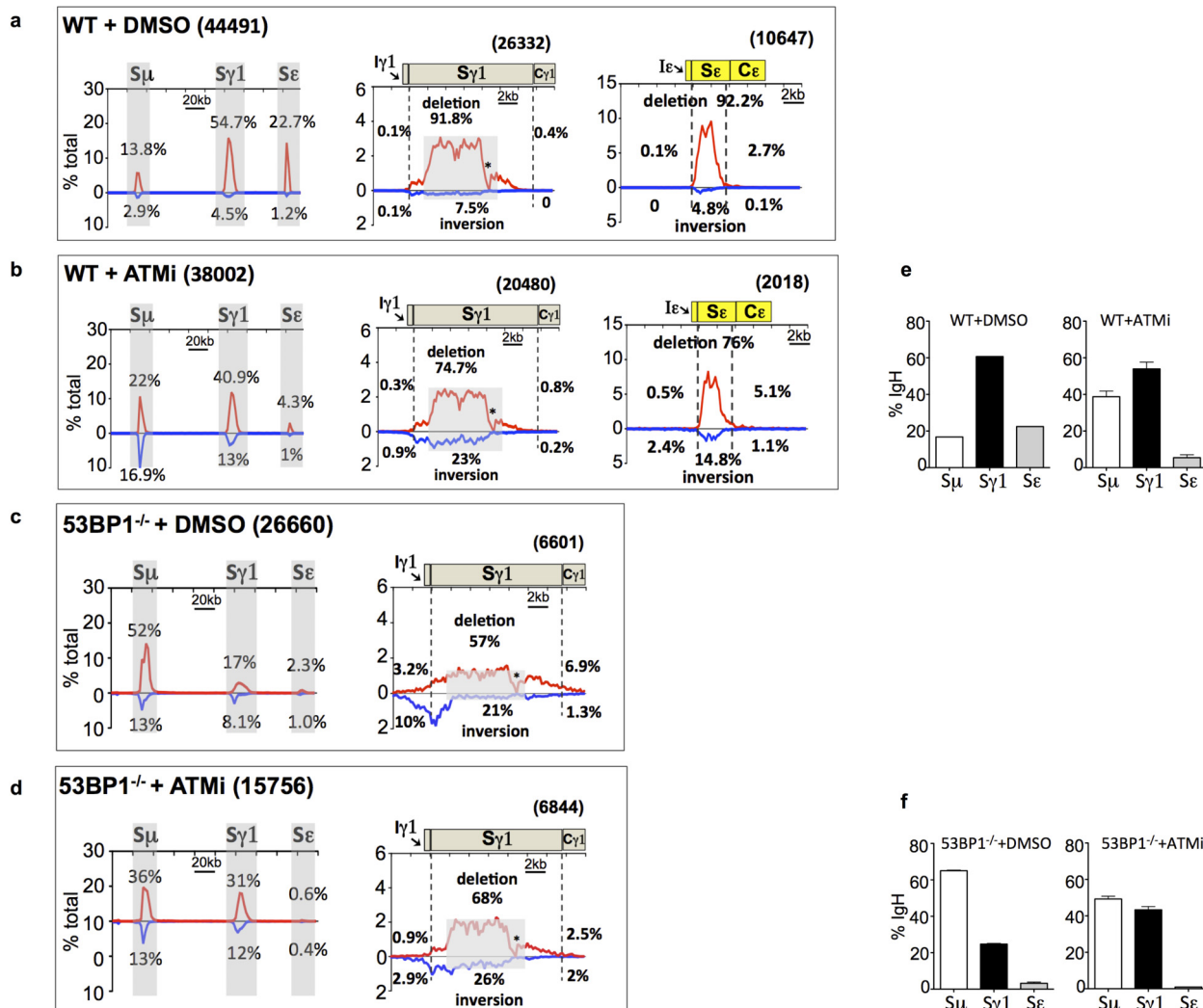
three independent libraries for each genotype. *P* values calculated by unpaired two-tailed *t*-tests. **f**, Bar graphs showing percentage of long resection of Sγ1 (left) and Sε (right) junctions in different genotypes as average \pm s.d. n.s., not significant. *P* values calculated by unpaired two-tailed *t*-tests. **g**, Bar graphs showing the number of junctions (average \pm s.d.) recovered from above experiments from 5'-broken end HTGTS libraries for the indicated genotypes (*n* = 3 for each) at the break site rSμ and downstream Sγ1 and Sε regions as a percentage of the total number of junctions mapped to the 200-kb *Igh* constant region. For detailed legends and further discussion refer to the Supplementary Information.



Extended Data Figure 8 | Orientation-biased joining of I-SceI DSBs at $S\gamma 1$ to AID-induced S-region breaks in various DSBR-deficient backgrounds.

a, Schematic illustration of the $\Delta S\gamma 1^{2xl}$ allele and joining outcomes from 3'-broken end (red arrow) to AID-initiated upstream $S\mu$ and downstream $S\epsilon$ DSBs. **b**, Linear distribution of junctions between $\Delta S\gamma 1^{2xl}$ 3'-broken end to AID-induced $S\mu/S\epsilon$ region DSBs in anti-CD40/IL4-stimulated wild-type (**b**, $n = 4$), $ATM^{-/-}$ (**c**, $n = 3$), $H2AX^{-/-}$ (**d**, $n = 3$) and $53BP1^{-/-}$ (**e**, $n = 4$) cells across the 200-kb *Igh* region (left panels), 10-kb $S\mu$ (middle panels) and $S\epsilon$ (right panels). **f**, Bar graphs (average \pm s.d.) showing the percentage of

junctions mapped at $\Delta S\gamma 1^{2xl}$ (break site), $S\mu$ and $S\epsilon$ over the total number of junctions in the 200-kb *Igh* constant region in different genotypes. **g**, Bar graphs (average \pm s.d.) showing the percentage of junctions from above various genotypes of $\Delta S\gamma 1^{2xl}$ 3'-broken end libraries mapped to $S\mu$ and $S\epsilon$ as average \pm s.d. **h**, Bar graph (average \pm s.d.) showing comparison of percentages of junctions cloned using $\Delta S\gamma 1^{2xl}$ 3'-broken end involving resection of $S\mu$ (top panel) and $S\epsilon$ (bottom panel) breaks in indicated backgrounds. For detailed legends and further discussion refer to the Supplementary Information.



Extended Data Figure 9 | Inhibition of resection in 53BP1-deficient B cells by an ATM inhibitor (ATMi) does not rescue directional CSR joining to Sγ1. **a–d**, Linear plots of pooled junctions across the 200-kb *Igh* constant region (left panels), the 20-kb region flanking downstream Sγ1 (middle panels) and Sε (right panels) from wild type plus DMSO control (**a**, $n = 2$), wild type plus ATMi (**b**, $n = 3$), 53BP1^{-/-} plus DMSO (**c**, $n = 3$) and 53BP1^{-/-} plus ATMi (**d**, $n = 3$) libraries are shown as above. **e**, Bar graph showing the

percentage of Sμ, Sγ1 and Sε junctions (average \pm s.d.) from wild type plus DMSO 5' Sμ libraries (left) and wild type plus ATMi 5'-broken end libraries (right, $n = 3$). **f**, Bar graph showing the percentage of Sμ, Sγ1 and Sε junctions (average \pm s.d.) from 53BP1^{-/-} plus DMSO (left) and 53BP1^{-/-} plus ATMi (right) 5' Sμ libraries. For detailed legends and further discussion refer to the Supplementary Information.

Extended Data Table 1 | Statistical comparison for orientation bias and resection in $S\gamma 1$ and $S\epsilon$ junctions from wild-type and DSBR-deficient B cell libraries

a	WT	$ATM^{-/-}$	$H2AX^{-/-}$	$53BP1^{-/-}$	$53BP1^{-/-} + Ai$	$Rif1^{-/-}$
WT (n=5)						
$ATM^{-/-}$ (n=3)	<0.0001					
$H2AX^{-/-}$ (n=3)	0.0057	0.0031				
$53BP1^{-/-}$ (n=8)	<0.0001	<0.0001	<0.0001			
$53BP1^{-/-} + Ai$ (n=3)	<0.0001	0.0039	0.001	0.1481		
$Rif1^{-/-}$ (n=3)	<0.0001	0.181	0.0094	<0.0001	0.0029	

b	WT	$ATM^{-/-}$	$H2AX^{-/-}$	$53BP1^{-/-}$	$53BP1^{-/-} + Ai$	$Rif1^{-/-}$
WT (n=5)						
$ATM^{-/-}$ (n=3)	<0.0001					
$H2AX^{-/-}$ (n=3)	0.0034	0.0291				
$53BP1^{-/-}$ (n=8)	<0.0001	0.0003	0.0002			
$53BP1^{-/-} + Ai$ (n=3)	N/A	N/A	N/A	N/A		
$Rif1^{-/-}$ (n=3)	0.0012	0.9274	0.1154	0.0003	N/A	

c	WT	$ATM^{-/-}$	$H2AX^{-/-}$	$53BP1^{-/-}$	$53BP1^{-/-} + Ai$	$Rif1^{-/-}$
WT (n=5)						
$ATM^{-/-}$ (n=3)	0.0002					
$H2AX^{-/-}$ (n=3)	0.0037	0.3012				
$53BP1^{-/-}$ (n=8)	<0.0001	0.0002	0.0008			
$53BP1^{-/-} + Ai$ (n=3)	<0.0001	0.158	0.7974	0.0005		
$Rif1^{-/-}$ (n=3)	N/A	N/A	N/A	N/A	N/A	

d	WT	$ATM^{-/-}$	$H2AX^{-/-}$	$53BP1^{-/-}$	$53BP1^{-/-} + Ai$	$Rif1^{-/-}$
WT (n=5)						
$ATM^{-/-}$ (n=3)	<0.0001					
$H2AX^{-/-}$ (n=3)	<0.0001	0.0279				
$53BP1^{-/-}$ (n=8)	<0.0001	0.0001	0.0003			
$53BP1^{-/-} + Ai$ (n=3)	N/A	N/A	N/A	N/A		
$Rif1^{-/-}$ (n=3)	<0.0001	0.0199	0.058	0.0062	N/A	

a, b, *P* values calculated by unpaired two-tailed *t*-test for the degree of bias in the $S\gamma 1$ (**a**) and $S\epsilon$ (**b**) junctions between wild-type and DSBR-deficient B cells with full-length $S\mu$ for experiments described in Figs 3 and 4. Orientation bias was calculated as described in the Methods for individual libraries. Numbers in parenthesis indicate independent experiments performed for each genotype. N/A, not available. **c, d**, *P* values calculated by unpaired two-tailed *t*-test for the level of resections in the $S\gamma 1$ (**c**) and $S\epsilon$ (**d**) junctions for experiments described above. Percentage of resection was calculated as described in the Methods for individual libraries. Numbers in parenthesis indicate independent experiments performed for each genotype. Ai, ATM inhibitor.

A four-helix bundle stores copper for methane oxidation

Nicolas Vita¹, Semeli Platsaki¹, Arnaud Baslé¹, Stephen J. Allen^{1†}, Neil G. Paterson², Andrew T. Crombie³, J. Colin Murrell³, Kevin J. Waldron¹ & Christopher Dennison¹

Methane-oxidizing bacteria (methanotrophs) require large quantities of copper for the membrane-bound (particulate) methane monooxygenase^{1,2}. Certain methanotrophs are also able to switch to using the iron-containing soluble methane monooxygenase to catalyse methane oxidation, with this switchover regulated by copper^{3,4}. Methane monooxygenases are nature's primary biological mechanism for suppressing atmospheric levels of methane, a potent greenhouse gas. Furthermore, methanotrophs and methane monooxygenases have enormous potential in bioremediation and for biotransformations producing bulk and fine chemicals, and in bioenergy, particularly considering increased methane availability from renewable sources and hydraulic fracturing of shale rock^{5,6}. Here we discover and characterize a novel copper storage protein (Csp1) from the methanotroph *Methylosinus trichosporium* OB3b that is exported from the cytosol, and stores copper for particulate methane monooxygenase. Csp1 is a tetramer of four-helix bundles with each monomer binding up to 13 Cu(I) ions in a previously unseen manner via mainly Cys residues that point into the core of the bundle. Csp1 is the first example of a protein that stores a metal within an established protein-folding motif. This work provides a detailed insight into how methanotrophs accumulate copper for the oxidation of methane. Understanding this process is essential if the wide-ranging biotechnological applications of methanotrophs are to be realized. Cytosolic homologues of Csp1 are present in diverse bacteria, thus challenging the dogma that such organisms do not use copper in this location.

Biology exploits copper to catalyse a range of important reactions, but use of this metal has been influenced by its availability and potential toxicity^{7–9}. In eukaryotes, excess copper is safely stored by cytosolic Cys-rich metallothioneins^{10–12}. Prokaryotic cytosolic copper-requiring

enzymes are not currently known, with most prokaryotes thought not to possess copper storage proteins. Copper-binding metallothioneins, such as those in eukaryotes, have been identified in pathogenic mycobacteria where they help detoxify Cu(I)¹³. Methanotrophs are Gram-negative bacteria that produce specialized membranes to harbour particulate methane monooxygenase (pMMO), which could either be discrete from or connected to the plasma membrane^{1,14,15}. These organisms have the typical machinery for copper efflux from the cytosol^{17,9}, but also possess the only currently characterized prokaryotic copper-uptake system; secreted modified peptides called methanobactins (mbtins)^{4,16}, which bind Cu(I) with high affinity^{17,18}, localize to the cytoplasm¹⁹, and are involved in soluble (s)MMO to pMMO switchover²⁰.

While searching for internalized Cu(I)-mbtin in the switchover methanotroph *M. trichosporium* OB3b, large amounts of soluble, protein-associated copper were detected. This was unexpected as bioinformatics predicts the transcriptional activator CueR as the only soluble copper protein²¹. To identify the copper-binding proteins in the most abundant copper pool, soluble extracts from cells grown under elevated copper were separated by anion-exchange and gel-filtration chromatography. Copper peaks match the intensity profiles of a protein band just below the 14.4 kDa marker (Fig. 1a, b), which has been purified to near homogeneity (Extended Data Fig. 1a, b and Fig. 1c). This band was excised (Fig. 1b, c) and identified by peptide mass fingerprinting as an uncharacterized conserved hypothetical protein (herein Csp1, Extended Data Fig. 1c). The mature protein (122 amino acids) contains 13 Cys residues with a predicted molecular mass of 12,591.4 Da, consistent with its migration on sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE) (Fig. 1).

Recombinant apo-Csp1 (12,589.8 Da) elutes from a gel filtration column with an apparent molecular mass of ~50 kDa (Fig. 2a), indicating

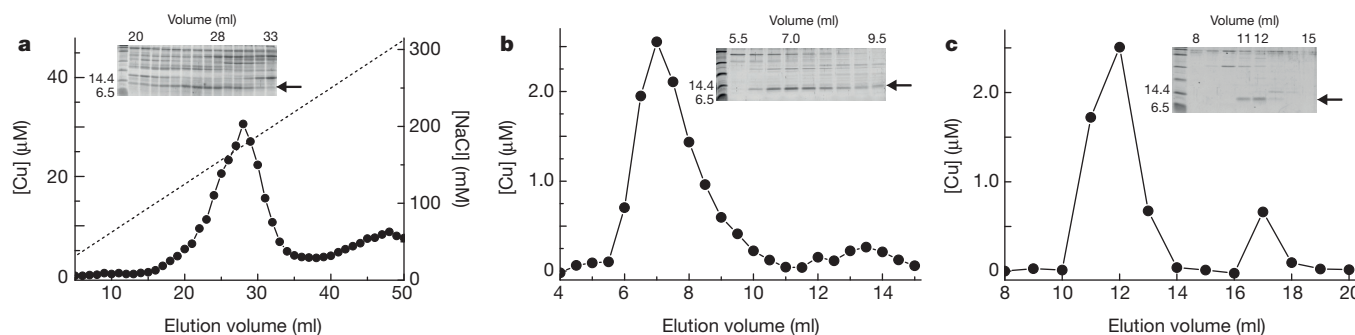


Figure 1 | Identification and purification of Csp1 from *M. trichosporium* OB3b. **a**, Copper content of anion-exchange fractions (NaCl gradient shown as a dashed line) of extract from *M. trichosporium* OB3b cells and the SDS-PAGE analysis of fractions 20–33. **b**, Copper content and SDS-PAGE analysis of the purification of the fraction containing the highest copper concentration (fraction 28) from **a** on a G100 gel-filtration column. A similar

anion-exchange fraction (Extended Data Fig. 1a) was purified on a Superdex 75 column (Extended Data Fig. 1b), with the copper content and SDS-PAGE analyses of eluted fractions shown in **c**. The band of interest that migrates below the 14.4 kDa marker is indicated in each panel with an arrow, and protein identification was performed on the bands from the 7.0 ml (**b**) and 12.0 ml (**c**) fractions.

¹Institute for Cell and Molecular Biosciences, Medical School, Newcastle University, Newcastle upon Tyne NE2 4HH, UK. ²Diamond Light Source, Harwell Science and Innovation Campus, Didcot OX11 0DE, UK. ³School of Environmental Sciences, University of East Anglia, Norwich Research Park, Norwich NR4 7TJ, UK. [†]Present address: Chemistry of Life Processes Institute, Northwestern University, 2170 Campus Drive, Evanston, Illinois 60208-2850, USA.

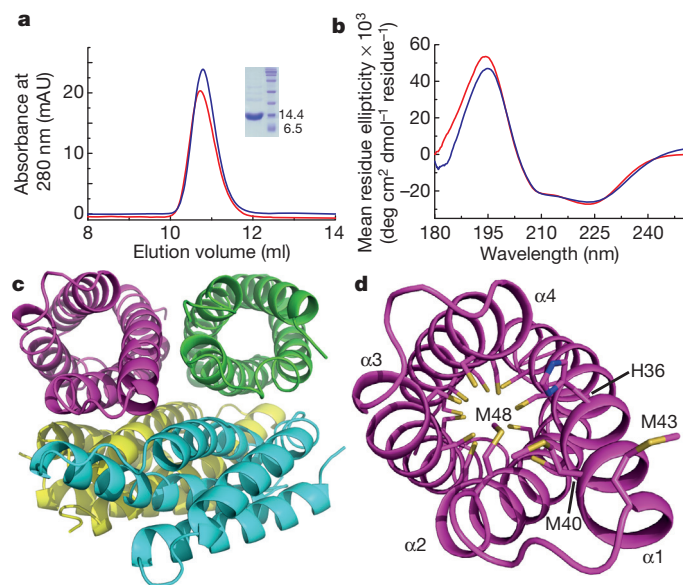


Figure 2 | The structure of apo-Csp1. **a**, Analytical gel-filtration chromatograms of apo-Csp1 (red line) and protein to which 14.0 molar equivalents of Cu(I) were added (blue line) for samples (100 μ M when injected) in 20 mM HEPES pH 7.5 containing 200 mM NaCl. The absorbance was monitored at 280 nm with the values for Cu(I)-Csp1 divided by 10 (see Extended Data Fig. 2a, b). Inset, SDS-PAGE analysis of the purified protein. **b**, Far-UV circular dichroism spectra of apo-Csp1 (red line) and Csp1 plus 14.0 equivalents of Cu(I) (blue line) at 39.6 and 35.7 μ M respectively in 100 mM phosphate pH 8.0. **c**, The tetrameric arrangement in the asymmetric unit of the crystal structure of apo-Csp1, with the side-chains of the Cys residues that point into the core of the four-helix bundle shown as sticks for one monomer in **d**. The opening into the core of the four-helix bundle is facing out in **d**, and involves His36, Met40, Met43 (on the extended α 1) and Met48.

a tetramer in solution (the native protein elutes at a similar volume (compare Fig. 2a and Extended Data Fig. 1b), demonstrating the same quaternary structure). The far-ultraviolet (UV) circular dichroism spectrum of Csp1 (Fig. 2b) reveals predominantly α -helical ($\sim 80\%$) secondary structure²². The asymmetric unit in the crystal structure of apo-Csp1 (Extended Data Table 1) consists of a tetramer of four-helix bundles ($\sim 75\%$ α -helix), involving two sets of adjacent monomers aligned in an anti-parallel manner, with pairs of monomers rotated by $\sim 55^\circ$ (Fig. 2c). The major contact areas between monomers are ~ 750 – 820 Å², consistent with the crystallographic tetramer being present in solution. All 13 Cys residues of apo-Csp1 point into the core of the four-helix bundle and none are involved in disulfide bonds (Fig. 2d). One end of the bundle, at which the amino (N) and carboxy (C) termini are found, is relatively hydrophobic while some Cys residues appear accessible at the opposite end of the molecule (Fig. 2d).

Csp1 is isolated from *M. trichosporium* OB3b with copper bound (Fig. 1 and Extended Data Fig. 1a), which, along with the arrangement of the Cys residues in the structure of the apo-protein (Fig. 2d), indicates a function in Cu(I) storage. To test this hypothesis, Cu(I) binding was analysed *in vitro* by monitoring the appearance of (S)Cys \rightarrow Cu(I) charge transfer bands (Extended Data Fig. 2a)^{10,23,24}, giving a stoichiometry of ~ 11 – 15 Cu(I) equivalents per Csp1 monomer (Extended Data Fig. 2b). In the presence of relatively low concentrations of the high-affinity chromophoric Cu(I) ligand bicinchoninic acid (BCA; $\log \beta_2 = 17.7$ (ref. 25)), apo-Csp1 binds all Cu(I) until ~ 12 – 14 equivalents have been added (see, for example, Fig. 3a). Furthermore, an apo-Csp1 sample incubated with ~ 25 equivalents of Cu(I) elutes from a gel-filtration column with ~ 12 – 13 equivalents of Cu(I) (Fig. 3b). The binding of Cu(I) has no significant effect on either the secondary or quaternary structure (Fig. 2a, b). The crystal structure of Cu(I)-Csp1 is shown in Fig. 3c (Extended Data Table 1). The anomalous difference density for data collected just below the

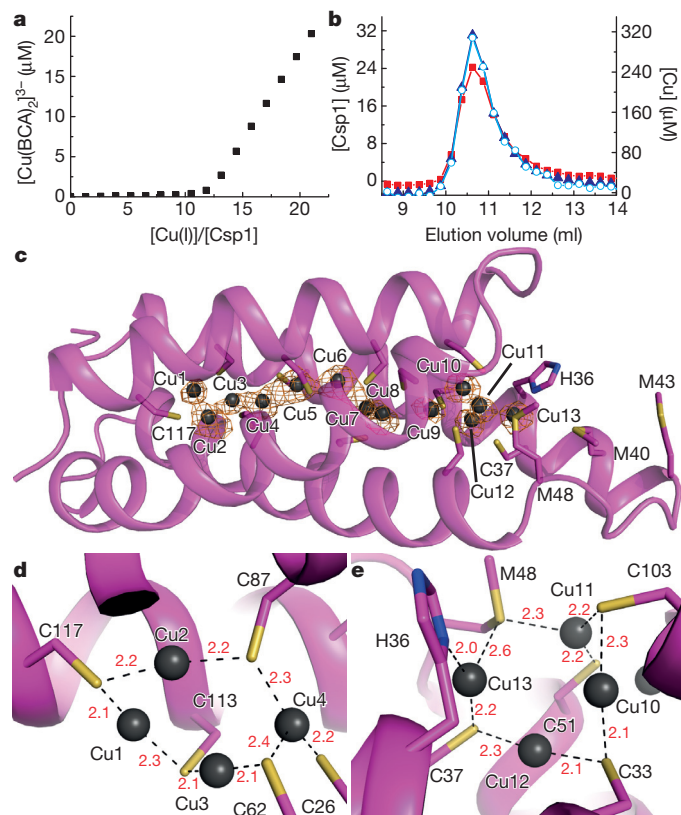


Figure 3 | Cu(I)-binding by Csp1. **a**, Plot of $[\text{Cu}(\text{BCA})_2]^{3-}$ concentration against the $[\text{Cu(I)}]/[\text{Csp1}]$ ratio upon titrating Cu(I) into apo-Csp1 (2.43 μ M) in the presence of 103 μ M BCA (same buffer as for Fig. 2a). $[\text{Cu}(\text{BCA})_2]^{3-}$ starts forming after ~ 12 equivalents of Cu(I) are added. **b**, Analytical gel-filtration chromatogram of Csp1 (116 μ M) mixed with ~ 25 equivalents of Cu(I) in the same buffer. Csp1 (Bradford, red squares), copper (atomic absorption spectroscopy, blue triangles) and Cu(I) (BCS in the presence of 7.6 M urea, open cyan circles) concentrations are shown. The main Csp1-containing fractions bind 11.8–12.9 equivalents of Cu(I). **c**, The structure of Cu(I)-Csp1 (chain A) including the anomalous difference density for copper contoured at 3.5σ (orange mesh). The copper ions (Cu1–Cu13 correspond to A1123–A1135 in the PDB file 5AJF) are represented as dark grey spheres, and the side chains of Cys, and other key residues, as sticks. The coordination of Cu(I) ions at the two ends of the four-helix bundle is shown in **d** and **e** with bond distances (in Å) in red.

copper-edge identifies 13 copper ions within the core of the four-helix bundle (Fig. 3c), bound predominantly by the 13 Cys residues. The oxidation state of copper in Csp1 crystals was analysed using X-ray absorption near-edge spectroscopy (Extended Data Fig. 2c). A well-defined peak at 8,984 eV, due to the Cu 1s \rightarrow 4p transition, is consistent with 2/3-coordinate Cu(I)²⁶.

The 13 Cu(I) ions are distributed throughout the core of the four-helix bundle of Csp1 (Fig. 3c–e). Ten of the Cu(I) sites involve bis-Cys (μ_2 -S for all the Cys ligands) ligation, with Cu(I)–S(Cys) bond lengths and S(Cys)–Cu–S(Cys) angles ranging from 2.0 to 2.3 Å and 142° to 177° , respectively (in chain A). Exceptions are Cu11 and Cu13 (Fig. 3e, see below), and Cu4 (Fig. 3d), which has three coordinating thiolates (S(Cys)–Cu–S(Cys) angles ranging from 90° to 145°). Ten Cu(I) ions are within 2.7 Å of a neighbouring metal (2.8 Å for Cu5 and Cu13, and 2.9 Å for Cu9) with some interacting with more than one Cu(I) (notably Cu7 and Cu10). Four of the Cu(I) sites (Cu1, Cu6, Cu8 and Cu12) are coordinated by the Cys residues from CXXXC motifs (Fig. 3c–e and Extended Data Figs 1c and 3), with the backbone carbonyl of the first Cys ligand close (2.1–2.3 Å) to the Cu(I) ion. At other Cu(I) sites the Cys ligands originate from adjacent α -helices (for example, Cu2, Cu3 and Cu4 in Fig. 3d and Cu10 in Fig. 3e). Cu11 involves ligation by two thiolates and Met48 (2.3 Å),

with bond angles ranging from 102° to 142° (Fig. 3e). Cu13 is also coordinated (Fig. 3e) by Met48 (2.6 Å), as well as by His36 (N^δ , 2.0 Å) and Cys37 (2.2 Å). These two atypical Cu(I) sites (Cu11 and Cu13) are found at the open end of the bundle, and with the nearby Met40 and Met43 (Fig. 3c) potentially help to recruit the metal.

Metal storage within an established protein-folding motif has not previously been observed. Iron is stored by ferritins using polymeric four-helix bundles, but with monomers forming a protein envelope that surrounds a ferric-oxide mineral core²⁷. Storing multiple Cu(I) ions within a four-helix bundle in Csp1 provides a stark contrast to unstructured apo-metallothioneins that fold around metal clusters. For example, a truncated form of yeast metallothionein binds a Cu(I)_8 -thiolate cluster using ten Cys residues with six three-coordinate and two two-coordinate sites¹¹. A four-helix bundle is formed upon Cu(I) addition to a synthetic peptide possessing a CXXC motif, and binds a Cu_4S_4 cluster²⁸. The arrangement of the Cu(I) ions within Csp1 is unprecedented in biology and inorganic chemistry.

Tetrameric Csp1 is capable of binding up to 52 Cu(I) ions, consistent with a role in copper storage. The major copper-requiring protein in *M. trichosporium* OB3b is pMMO. Regardless of the uncertainty about the structure of the specialized membranes that house pMMO, cytosolic copper must cross a membrane before incorporation into this enzyme. Csp1 and its closely related homologue Csp2 possess signal peptides (Extended Data Figs 1c and 3), predicted as targeting the twin arginine translocation (Tat) machinery²⁹, and therefore locate outside the cytosol. To test whether Csp1 and Csp2 store copper for pMMO, the $\Delta\text{csp1/csp2}$ double mutant strain of *M. trichosporium* OB3b was constructed. Switchover to sMMO for cells transferred from high to low copper is significantly faster in $\Delta\text{csp1/csp2}$ than in the wild-type strain, and sMMO activity is 1.8 times greater in the former after almost 28 h (Extended Data Fig. 4). These data are not inconsistent with Csp1 and Csp2 storing, and potentially also chaperoning, copper for pMMO, thus allowing *M. trichosporium* OB3b to use this enzyme longer for growth on methane when copper becomes limiting, although this hypothesis has not been explicitly tested.

An important attribute of a metal storage protein is its metal affinity. Upon increasing the concentration of BCA, Cu(I) starts to be withheld from Csp1 (Fig. 4a and Extended Data Fig. 5a–f). The buffering of free Cu(I) by ligands such as BCA and the tighter chromophoric Cu(I) ligand bathocuproine disulfonate (BCS; $\log\beta_2 = 20.8$ (ref. 25), see Extended Data Fig. 6a, b) has been used to obtain an average Cu(I) affinity for Csp1 of $\sim 1 \times 10^{17} \text{ M}^{-1}$ (Fig. 4b, c and Extended Data Fig. 6c, d). Mbtin, the copper-chelating ligand produced by *M. trichosporium* OB3b, has a much tighter Cu(I) affinity¹⁷, and stoichiometrically removes Cu(I) from Csp1 in ~ 1 h (Fig. 4d and Extended Data Fig. 7a–d). This high affinity makes copper removal from imported mbtin potentially problematical (for example apo-Csp1 cannot directly acquire Cu(I) from mbtin; Extended Data Fig. 7e), and Cu(I)-mbtin may need to be degraded within a cell to release its metal cargo¹⁸. Csp1 is present at elevated copper levels in *M. trichosporium* OB3b and sequesters the metal (Fig. 1 and Extended Data Fig. 1a), while mbtin production is suppressed under these conditions^{4,16}. As copper levels in the cell decrease, mbtin will be produced and could play a role in removing and using Csp1-bound copper.

Another Csp1 homologue, Csp3, is also encoded in the *M. trichosporium* OB3b genome (Extended Data Fig. 3), and is widespread in bacteria (Extended Data Fig. 8). This includes the functionally uncharacterized proteins from *Nitrospira multiformis* and *Pseudomonas aeruginosa* (apo-structures are available in Protein Data Bank (PDB) under accession numbers 3LMF and 3KAW) that also have all Cys residues pointing to the core of four-helix bundles. Csp3 has no signal peptide (Extended Data Fig. 3) and therefore, unlike Csp1 and Csp2, is presumably cytosolic. If Csp1 and Csp2 are exported via the Tat system they would fold in the cytosol²⁹, perhaps to prevent disulfide formation. Tat-export might also imply copper acquisition before export²⁹, allowing transport of large amounts of copper away from systems that

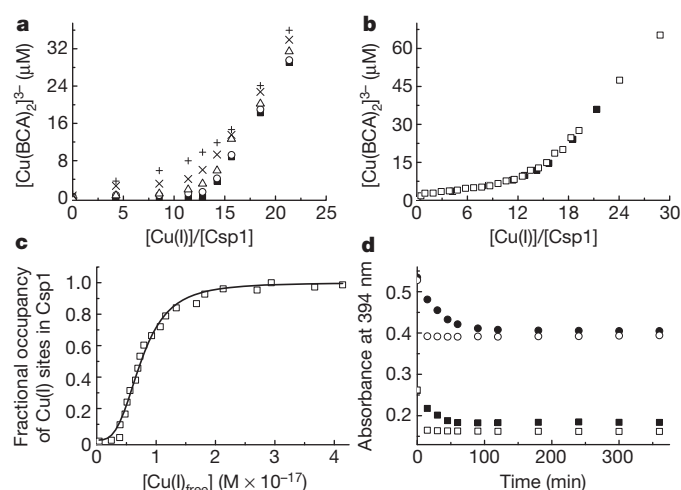


Figure 4 | Cu(I) affinity of Csp1 and Cu(I) release. **a**, Plots of $[\text{Cu(BCA)}_2]^{3-}$ concentration against the $[\text{Cu(I)}]/[\text{Csp1}]$ ratio for mixtures of apo-Csp1 (3.57 μM) and Cu(I) in the presence of 120 μM (filled squares), 300 μM (open circles), 600 μM (open triangles), 900 μM (cross) and 1,200 μM ('+' symbol) BCA (all data acquired after incubation for 41 h). **b**, Plot of $[\text{Cu(BCA)}_2]^{3-}$ concentration against the $[\text{Cu(I)}]/[\text{Csp1}]$ ratio for mixtures of apo-Csp1 (3.61 μM) and Cu(I) in the presence of 1,210 μM BCA (open squares) for 20 h, along with the data from **a** at 1,200 μM BCA (filled squares). **c**, Fractional occupancy of Cu(I)-binding sites in Csp1 (maximum value 11.7 equivalents) at different concentrations of free Cu(I) from the data shown in **b** at 1,210 μM BCA up to a $[\text{Cu(I)}]/[\text{Csp1}]$ ratio of 19.2. The solid line shows the fit of the data to the nonlinear Hill equation giving an average dissociation constant for Cu(I), K_{Cu} of $(7.5 \pm 0.1) \times 10^{-18} \text{ M}$ ($n = 3.1 \pm 0.2$, see Extended Data Fig. 6c, d). **d**, Plots of the absorbance at 394 nm (spectra in Extended Data Fig. 7a–d) against time after the addition of Cu(I)-Csp1 (1.02 μM) loaded with 13.0 equivalents of Cu(I) to 13.4 μM (filled squares) and 27.4 μM (filled circles) apo-mbtin, and Cu(I) (13.3 μM) to 13.4 μM (open squares) and 27.1 μM (open circles) apo-mbtin. All experiments were performed in the same buffer as for Fig. 2a.

remove this metal from the cytosol (CueR and copper-transporting ATPases) and into the same compartment housing pMMO. Csp1 (and Csp2) can store large quantities of copper for, and potentially deliver the metal to, pMMO, an enzyme of great environmental importance that has tremendous biotechnological potential for the utilization and mitigation of methane. The prediction would be that Csp3 can store copper in the cytosol, not only in *M. trichosporium* OB3b but in many other bacteria. This raises the possibility that there are cytosolic copper-requiring enzymes in bacteria still to be discovered.

Online Content Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 22 December 2014; accepted 23 June 2015.

Published online 26 August 2015.

- Hanson, R. S. & Hanson, T. E. Methanotrophic bacteria. *Microbiol. Rev.* **60**, 439–471 (1996).
- Balasubramanian, R. *et al.* Oxidation of methane by a biological dicopper centre. *Nature* **465**, 115–119 (2010).
- Murrell, J. C., McDonald, I. R. & Gilbert, B. Regulation of expression of methane monooxygenases by copper ions. *Trends Microbiol.* **8**, 221–225 (2000).
- Hakemian, A. S. & Rosenzweig, A. C. The biochemistry of methane oxidation. *Annu. Rev. Biochem.* **76**, 223–241 (2007).
- Jiang, H. *et al.* Methanotrophs: multifunctional bacteria with promising applications in environmental bioengineering. *Biochem. Eng. J.* **49**, 277–288 (2010).
- Haynes, C. A. & Gonzalez, R. Rethinking biological activation of methane and conversion to liquid fuels. *Nature Chem. Biol.* **10**, 331–339 (2014).
- Ma, Z., Jacobsen, F. E. & Giedroc, D. P. Coordination chemistry of bacterial metal transport and sensing. *Chem. Rev.* **109**, 4644–4681 (2009).
- Festa, R. A. & Thiele, D. J. Copper: an essential metal in biology. *Curr. Biol.* **21**, R877–R883 (2011).

9. Argüello, J. M., Raimunda, D. & Padilla-Benavides, T. Mechanisms of copper homeostasis in bacteria. *Front. Cell. Infect. Microbiol.* **3**, 73 (2013).
10. Poutney, D. L., Schauwecker, I., Zarn, J. & Vašák, M. Formation of mammalian Cu₂-metallothionein *in vitro*: evidence for the existence of two Cu(I)₄-thiolate clusters. *Biochemistry* **33**, 9699–9705 (1994).
11. Calderone, V. *et al.* The crystal structure of yeast copper thionein: the solution of a long-lasting enigma. *Proc. Natl Acad. Sci. USA* **102**, 51–56 (2005).
12. Sutherland, D. E. K. & Stillman, M. J. The “magic numbers” of metallothionein. *Metallomics* **3**, 444–463 (2011).
13. Gold, B. *et al.* Identification of a copper-binding metallothionein in pathogenic bacteria. *Nature Chem. Biol.* **4**, 609–616 (2008).
14. Davies, S. L. & Whittenbury, R. Fine structure of methane and other hydrocarbon-utilising bacteria. *J. Gen. Microbiol.* **61**, 227–232 (1970).
15. Reed, W. M., Titus, J. A., Dugan, P. R. & Pfister, R. M. Structure of *Methylosinus trichosporium* exospores. *J. Bacteriol.* **141**, 908–913 (1980).
16. Kim, H. J. *et al.* Methanobactin, a copper-acquisition compound from methane-oxidizing bacteria. *Science* **305**, 1612–1615 (2004).
17. El Ghazouani, A. *et al.* Copper-binding properties and structures of methanobactins from *Methylosinus trichosporium* OB3b. *Inorg. Chem.* **50**, 1378–1391 (2011).
18. El Ghazouani, A. *et al.* Variations in methanobactin structure influence copper utilization by methane-oxidizing bacteria. *Proc. Natl Acad. Sci. USA* **109**, 8400–8404 (2012).
19. Balasubramanian, R., Kenney, G. E. & Rosenzweig, A. C. Dual pathways for copper uptake by methanotrophic bacteria. *J. Biol. Chem.* **286**, 37313–37319 (2011).
20. Semrau, J. D. *et al.* Methanobactin and MmoD work in concert to act as the ‘copper-switch’ in methanotrophs. *Environ. Microbiol.* **15**, 3077–3086 (2013).
21. Stein, L. Y. *et al.* Genome sequence of the obligate methanotroph *Methylocystis trichosporium* strain OB3b. *J. Bacteriol.* **192**, 6497–6498 (2010).
22. Barrow, C. J., Yasuda, A., Kenny, P. T. M. & Zagorski, M. G. Solution conformations and aggregational properties of synthetic amyloid β -peptides of Alzheimer’s disease. Analysis of circular dichroism spectra. *J. Mol. Biol.* **225**, 1075–1093 (1992).
23. Cobine, P. A. *et al.* Copper transfer from the Cu(I) chaperone, CopZ, to the repressor, Zn(II)CopY: metal coordination environments and protein interactions. *Biochemistry* **41**, 5822–5829 (2002).
24. Badarau, A., Firbank, S. J., McCarthy, A. A., Banfield, M. J. & Dennison, C. Visualizing the metal-binding versatility of copper trafficking sites. *Biochemistry* **49**, 7798–7810 (2010).
25. Bagchi, P., Morgan, M. T., Bacsa, J. & Fahrni, C. J. Robust affinity standards for Cu(I) biochemistry. *J. Am. Chem. Soc.* **135**, 18549–18559 (2013).
26. Kau, L. S., Spira-Solomon, D. J., Penner-Han, J. E., Hodgson, K. O. & Solomon, E. I. X-ray absorption edge determination of the oxidation state and coordination number of copper: applications to the type 3 site in *Rhus vernicifera* laccase and its reaction with oxygen. *J. Am. Chem. Soc.* **109**, 6433–6442 (1987).
27. Theil, E. C. Ferritin protein nanocages use ion channels, catalytic sites, and nucleation channels to manage iron/oxygen chemistry. *Curr. Opin. Chem. Biol.* **15**, 304–311 (2011).
28. Kharenko, O. A., Kennedy, D. C., Demeler, B., Maroney, M. J. & Ogawa, M. Y. Cu(I) luminescence from the tetranuclear Cu₄S₄ cofactor of a synthetic 4-helix bundle. *J. Am. Chem. Soc.* **127**, 7678–7679 (2005).
29. Palmer, T. & Berks, B. C. Moving folded proteins across the bacterial cell membrane. *Microbiology* **149**, 547–556 (2003).

Acknowledgements We thank staff of the Diamond Light Source for help with diffraction data collection, J. Gray for performing mass spectrometry studies, the School of Civil Engineering and Geosciences and D. Graham for access to facilities at the very start of this work, A. Badarau for discussions about determining Cu(I) affinities and S. J. Firbank for structural modelling at the initial stages of this project. This work was supported by Biotechnology and Biological Sciences Research Council (grant BB/K008439/1 to C.D. and K.J.W.) and Newcastle University (part funding of a PhD studentship for S.P.). K.J.W. was supported by a Sir Henry Dale Fellowship funded by the Wellcome Trust and the Royal Society (098375/Z/12/Z).

Author Contributions C.D. and K.J.W. designed the research, N.V. and S.J.A. performed *in vitro* characterization of Csp1 and Cu(I) binding, S.P. isolated Csp1 from *M. trichosporium* OB3b and crystallized the protein, A.B. and N.G.P. solved the crystal structures, and A.T.C. and J.C.M. constructed and characterized the Δ csp1csp2 *M. trichosporium* OB3b strain. C.D. wrote the paper with help from all authors.

Author Information Atomic coordinates have been deposited in the Protein Data Bank under accession numbers 5AJE and 5AJF for apo- and Cu(I)-Csp1 respectively. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to C.D. (christopher.dennison@ncl.ac.uk).

METHODS

Identification and purification of copper proteins from *M. trichosporium* OB3b. *M. trichosporium* OB3b cultures were grown as described¹⁷ at 27 °C in a 5 l fermentor agitated at 250 r.p.m. in nitrate minimal salts (NMS) medium supplemented with 10 µM iron and typically 5 µM copper. Cultures were analysed for sMMO activity as described¹⁷. Cells harvested at an absorbance, $A_{600\text{ nm}}$, typically between 1.1 and 2.2 were collected by centrifugation (4 °C) at 9,000g and pellets washed with 20 mM HEPES pH 8.8 followed by the same buffer containing 10 mM EDTA. The cell pellet was resuspended in 20 mM HEPES pH 8.8 and lysed by freeze grinding in liquid nitrogen³⁰. The lysate was allowed to thaw in an anaerobic chamber (Belle Technology, $[O_2] \ll 2$ ppm), before loading into ultracentrifuge tubes sealed in the anaerobic chamber and centrifuged at 160,000g for 1 h at 10 °C. The supernatant was recovered inside the anaerobic chamber, diluted fivefold and loaded (either 59 or 90 mg protein from 10 and 161 of cells, respectively) onto a 5 ml HiTrap Q HP anion-exchange column (GE Healthcare). For the purification of extracts from 101 of cells, the HiTrap column was eluted with a linear NaCl gradient (0–500 mM) inside the anaerobic chamber with a homemade mixing device (total volume 80 ml). For the 16 l preparation, the HiTrap column was eluted on an Akta Purifier with a linear NaCl gradient (0–250 mM) using thoroughly degassed and nitrogen-purged buffers (total volume 80 ml). Copper content in eluted fractions (1 ml) was measured by inductively coupled plasma mass spectrometry (Thermo Electron, X series). Samples were diluted tenfold in 2% nitric acid containing 20 µg l⁻¹ silver as internal standard, and analysed for ⁶³Cu and ¹⁰⁷Ag in standard mode (100 reads, 30 ms dwell, 3 channels, 0.02 atomic mass unit separation, in triplicate). Copper concentration was determined by comparison with matrix-matched elemental standard solutions. Copper-containing fractions were analysed by SDS-PAGE using Oriole fluorescent gel stain (Bio-Rad). All images of fluorescently labelled gels were inverted to make bands clearer in print.

Gel-filtration chromatography of copper-containing fractions was performed on either a Sephadex G100 (Sigma) packed column (1 cm × 20 cm) inside the anaerobic chamber, or on a Superdex S75 10/300 GL (GE Healthcare) column in 20 mM HEPES pH 7.5 plus 200 mM NaCl (thoroughly degassed and nitrogen-purged for the Superdex 75 column that was attached to an Akta Purifier) and at flow rates of 0.35 and 0.8 ml min⁻¹ respectively (fraction size 1 ml). Proteins whose intensity on SDS-PAGE gels correlated with copper concentration profiles were excised from gels and underwent peptide mass fingerprinting³¹. Digestion with trypsin was performed at an E:S ratio of 1:100 overnight in 50 mM NH₄HCO₃ pH 8. The resultant peptides were resuspended in 0.1% aqueous trifluoroacetic acid and desalted using C18 ZipTips (Millipore). Peptides were then separated on a NanoAcquity liquid chromatography system (Waters) using a 75 µm × 100 mm C18 capillary column (Waters). A linear gradient from 1 to 50% acetonitrile in 0.1% aqueous formic acid was applied over 30 min at a flow rate of 0.3 ml min⁻¹. Eluted peptides were detected using a linear trap quadrupole Fourier transform mass spectrometer (ThermoElectron) in positive ionization mode with scans over 300–1,500 *m/z* in data-dependent mode and a Fourier transform mass spectrometry (MS) resolution setting of 50,000. The top five ions in the parent scan were subjected to MS/MS analysis in the linear ion trap region. The proteins from which detected peptides originated were identified using the Mascot MS/MS ion search tool (Matrix Science) by comparison against the entire database of proteobacteria in NCBI. SignalP and TatP were used to identify putative signal sequences³².

Cloning the *csp1* gene. The *csp1* gene without its predicted signal peptide (that is, Gly25 to Ala144, see Extended Data Fig. 1c) was amplified from *M. trichosporium* OB3b genomic DNA using primers Csp1_F and Csp1_R (Extended Data Table 2) and cloned into pGEMT, which introduced a Met residue at the N terminus. Both strands of the gene were verified by sequencing, which was subsequently cloned into the NdeI and NcoI sites of pET29a (pET29a_Csp1).

Expression and purification of Csp1. *Escherichia coli* BL21 (DE3) transformed with pET29a_Csp1 was grown in LB media at 37 °C (100 µg ml⁻¹ kanamycin) until an $A_{600\text{ nm}}$ of ~0.6. Cells were induced with 1 mM isopropyl β-D-thiogalactopyranoside, harvested after 6 h and stored at -20 °C. Pellets were resuspended in 20 mM Tris pH 8.5 plus 1 mM dithiothreitol (DTT), sonicated and centrifuged at 40,000g for 30 min. The supernatant was diluted fivefold in 20 mM Tris pH 8.5 containing 1 mM DTT and loaded onto a HiTrap Q HP column (5 ml) equilibrated in the same buffer. Proteins were eluted with a linear NaCl gradient (0–300 mM, total volume ~200 ml). Csp1-containing fractions, identified by SDS-PAGE, were combined, diluted tenfold in 10 mM Tris pH 7.5 plus 1 mM DTT, applied to a HiTrap Q HP column (5 ml) equilibrated in the same buffer and eluted using a linear NaCl gradient (0–200 mM, total volume ~200 ml). The purest fractions, identified by SDS-PAGE, were combined and thoroughly exchanged into 20 mM HEPES pH 7.5 plus 200 mM NaCl using either a stirred cell or a centrifugal filter unit (typically with 10 kDa molecular mass cut-off membranes). Except for crystallization of the apo-protein, Csp1 was further purified by gel-filtration chromatography

on a Superdex 75 10/300 GL column (GE Healthcare) equilibrated with 20 mM HEPES pH 7.5 containing 200 mM NaCl. The protein was found to contain no copper and zinc using atomic absorption spectrometry (AAS, with an M Series spectrometer, Thermo Electron) typically with ten standards containing up to 1.8 ppm copper and 1.0 ppm zinc in 2% HNO₃ using the standard calibration method. The mass of purified Csp1 was verified both by matrix assisted laser desorption/ionization time-of-flight and Fourier transform ion cyclotron resonance MS. The Met residue introduced at the N terminus during cloning is largely cleaved in the overexpression host, giving a purified protein with an experimental mass of 12,589.8 Da, consistent with the calculated value (12,591.4 Da) for a mature protein having Gly1 and Ala122 at the N and C termini respectively.

Isolation, purification and quantification of mbtin. The apo- and Cu(I)-forms of full-length mbtin from *M. trichosporium* OB3b was isolated, purified and quantified as described previously¹⁷.

Quantification of Csp1. Apo-Csp1 was quantified by the reduction of 5,5'-dithiobis(2-nitrobenzoic acid) (DTNB, Ellman's reagent) in a sealed anaerobic quartz cuvette monitored at 412 nm using a λ35 UV-visible (UV-vis) spectrophotometer (Perkin Elmer)³³. The reaction was initiated in the anaerobic chamber by the addition of apo-Csp1 (final concentration typically 0.2–4 µM) to a buffered solution of DTNB (240–480 µM) in the presence of urea (final concentration >7 M). Under these conditions Csp1 rapidly unfolds, particularly at concentrations <8 µM (Extended Data Fig. 5g, h). The buffer was typically 20 mM HEPES pH 7.5 plus 200 mM NaCl and 1 mM EDTA, but in some cases 100 mM phosphate pH 8.0 plus 1 mM EDTA was used (the 5–10 mM DTNB stock solution was always made in 100 mM phosphate pH 8.0 plus 1 mM EDTA). After approximately 10–20 min, the absorbance at 412 nm reached a plateau and it was assumed that all 13 Cys residues of unfolded apo-Csp1 had reacted with DTNB (in the absence of urea very little reaction with DTNB occurs, consistent with the structure of the apo-protein (Fig. 2d)). Apo-Csp1 (~4–23 µM) incubated overnight in the anaerobic chamber with DTT (~2–6.4 mM) and desalted on a PD10 column (GE Healthcare) was also quantified with DTNB in 20 mM HEPES pH 7.5 plus 200 mM NaCl, 1 mM EDTA and >7 M urea (again little reaction with DTNB occurred in the absence of urea). At higher DTT concentrations, samples were desalted twice to ensure all DTT had been removed. For the DTNB reaction, a molar absorption coefficient (ϵ value) of 14,150 M⁻¹ cm⁻¹ at 412 nm (refs 24, 33, 34) was verified using glutathione with and without >7 M urea in both 20 mM HEPES pH 7.5 plus 200 mM NaCl and 1 mM EDTA (three times), and in 100 mM phosphate pH 8.0 plus 1 mM EDTA (twice). Apo-Csp1 concentrations were also determined using the Bradford assay (Coomassie Plus protein assay kit, Thermo Scientific) with BSA standards (0–1,000 µg ml⁻¹). The ratio of apo-Csp1 concentration obtained using the Bradford and DTNB assays (Bradford:DTNB ratio) was 1.11 ± 0.12 (*n* = 27) for samples not treated with DTT and 1.03 ± 0.04 (*n* = 9) for samples that were reduced before these assays. Incubation with DTT does not have a significant effect on the thiol count and this step was excluded from all subsequent experiments as contamination of apo-Csp1 with even trace amounts of DTT would influence quantification using DTNB.

Investigating Cu(I)-binding to Csp1. Cu(I) stock solutions (50–100 mM [Cu(CH₃CN)₄]PF₆ in 100% acetonitrile) were diluted to ~1–12 mM in 20 mM HEPES pH 7.5 plus 200 mM NaCl in the anaerobic chamber^{24,35}. Copper concentrations were determined by AAS, and Cu(I) was quantified anaerobically by UV-vis with the chromophoric high-affinity Cu(I) ligands BCS and BCA by monitoring formation of [Cu(BCS)₂]³⁻ and [Cu(BCA)₂]³⁻ at 483 nm (ϵ = 12,500 M⁻¹ cm⁻¹) and 562 nm (ϵ = 7,700 M⁻¹ cm⁻¹) respectively^{35–38}. Cu(I) was added to apo-Csp1 by mixing the appropriate amount of the buffered Cu(I) solution with apo-protein (~2–200 µM) that had been quantified by DTNB, in 20 mM HEPES pH 7.5 plus 200 mM NaCl in the anaerobic chamber. Using DTNB in urea is a more precise method for quantifying apo-Csp1 than the Bradford assay, particularly at low protein concentrations, and was therefore used routinely. However, the DTNB assay in urea cannot be used for Cu(I)-Csp1 owing to very slow unfolding (Extended Data Fig. 5i). Therefore, for most Cu(I)-Csp1 samples the number of Cu(I) equivalents quoted are based on the apo-protein concentration determined by DTNB (the Cu(I)-Csp1 concentrations take into account dilution of the sample due to the addition of Cu(I), and the number of Cu(I) equivalents quoted are those in the final sample). The [Cu(I)]/[Csp1] ratio was routinely checked using protein (Bradford assay) and copper (AAS and with 2.5 mM BCS both in the absence and presence of >7 M urea, compared in Extended Data Fig. 2d) quantifications, with good agreement. For titrations (performed more than ten times), Cu(I) from the buffered solution was added to apo-Csp1 (~2–20 µM) in 20 mM HEPES pH 7.5 plus 200 mM NaCl in a sealed anaerobic quartz cuvette using a gastight syringe (Hamilton). The immediate appearance of ligand-to-metal charge transfer (LMCT) bands, characteristic of Cu(I) coordination by thiolates^{10,23,24}, was monitored in the UV region. Emission spectra were acquired on a Cary Eclipse fluorescence spectrophotometer (Varian) by exciting at 280 nm and following the emission in the 400–700 nm range using emission and

excitation slits of 20 and 10 nm respectively. The concentration of the Cu(I) solution was regularly checked during titrations, usually with BCS, and replaced as required. **Competition between Csp1 and chromophoric ligands.** The binding of Cu(I) by Csp1 in the presence of either BCA or BCS was investigated in a variety of ways. Additions of Cu(I) to Csp1 (~1.6–3.0 μM) were performed in the presence of ~90–110 μM BCA in 20 mM MES at pH 5.5 (the $\text{p}K_{\text{a}}$ of BCA is 3.8 (ref. 25) and its β_2 value is therefore hardly affected at this pH value) and 6.5, HEPES at pH 7.5, TAPS at pH 8.5 and CHES at pH 9.5, all plus 200 mM NaCl. $[\text{Cu}(\text{BCA})_2]^{3-}$ concentrations were determined under anaerobic conditions as described above. This experiment was repeated twice at pH 6.5 and at least three and up to six times at other pH values, except at pH 7.5, when it was performed more than ten times. Apart from at pH 5.5, equilibration typically took less than 10 min (~20 min at ~11–15 Cu(I) equivalents). During these titrations the concentration of the Cu(I) solution was regularly checked and replaced as required. For experiments at pH 5.5 (MES), and at higher concentrations of BCA (up to ~1.2 mM and using ~2.0–3.7 μM apo-Csp1), UV-vis spectra were acquired between 4 and 48 h after mixing, to ensure equilibration had occurred (experiments at pH 6.5 and 9.5 were repeated two and four times respectively, while the experiment at pH 7.5 was repeated six times). The final pH values of samples were checked at the end of experiments and were within 0.1 pH units (0.2 at pH 5.5) of the buffer used.

A comparison of the ability of Csp1 to compete with BCA and BCS was performed at least three times by incubating Cu(I)-Csp1 (~2.4–2.7 μM) loaded with ~10–13 equivalents of Cu(I) with various concentrations of either ligand in 20 mM HEPES pH 7.5 plus 200 mM NaCl in the anaerobic chamber. $[\text{Cu}(\text{BCS})_2]^{3-}$ and $[\text{Cu}(\text{BCA})_2]^{3-}$ concentrations were determined by UV-vis under anaerobic conditions of mixtures incubated for various times (4–48 h) with very little change. Furthermore, apo-Csp1 (~2.5–2.8 μM) plus BCS (~100 or 250 μM) was incubated anaerobically with 0 to ~22 Cu(I) equivalents in 20 mM HEPES pH 7.5 plus 200 mM NaCl (repeated four times). The absorbance at 483 nm was monitored anaerobically after 4 h and up to 43 h after mixing (no variation observed). The kinetics of Cu(I) removal from Cu(I)-Csp1 (~0.3–1.6 μM) loaded with ~11–14 equivalents of Cu(I) by ~2,500 μM BCS was compared (five times) in the absence and presence of >7 M urea monitored anaerobically at 483 nm in 20 mM HEPES pH 7.5 plus 200 mM NaCl.

Estimation of the average Cu(I) affinity of Csp1. The average Cu(I) affinity of Csp1 was estimated by determining the Cu(I) occupancy of Csp1 as a function of the concentration of free Cu(I) ($[\text{Cu}(\text{I})_{\text{free}}]$) buffered using either BCA or BCS^{39,40}. Apo-Csp1 (~2.7 and ~3.6 μM respectively) in 20 mM HEPES pH 7.5 plus 200 mM NaCl was mixed anaerobically with increasing Cu(I) concentrations in the presence of BCS (101 μM) or BCA (1,210 and 2,000 μM). Mixtures were incubated for up to 67 h. The final pH values of samples were checked at the end of experiments. The concentrations of $[\text{Cu}(\text{BCA})_2]^{3-}$ and $[\text{Cu}(\text{BCS})_2]^{3-}$ ($[\text{Cu}(\text{L})_2]$, where L = BCA or BCS) were determined as described above and the concentration of Cu(I) bound to Csp1 ($[\text{Cu}(\text{I})_{\text{Csp1}}]$) was calculated using equation (1), where $[\text{Cu}(\text{I})_{\text{total}}]$ is the total concentration of Cu(I) added:

$$[\text{Cu}(\text{I})_{\text{Csp1}}] = [\text{Cu}(\text{I})_{\text{total}}] - [\text{Cu}(\text{L})_2] \quad (1)$$

$[\text{Cu}(\text{I})_{\text{free}}]$ was calculated using equation (2):

$$[\text{Cu}(\text{I})_{\text{free}}] = \frac{[\text{Cu}(\text{L})_2]}{[\text{L}]^2 \times \beta_2} \quad (2)$$

where $[\text{L}]$ is the total free ligand concentration ($[\text{L}] = [\text{L}] - 2[\text{Cu}(\text{L})_2]$) and β_2 is the affinity of either BCA or BCS for Cu(I) ($\log \beta_2 = 17.7$ and 20.8 respectively²⁵). The Cu(I) occupancy was determined by dividing $[\text{Cu}(\text{I})_{\text{Csp1}}]$ obtained from equation (1) by the Csp1 concentration. The fractional Cu(I) occupancy was calculated using the maximum value observed in a particular experiment (see below) and plots against $[\text{Cu}(\text{I})_{\text{free}}]$ were fitted to the nonlinear form of the Hill equation (3) to obtain the average dissociation constant of Csp1 for Cu(I) (K_{Cu}) and Hill coefficient (n value):

$$\text{Fractional Cu (I) occupancy} = \frac{[\text{Cu}(\text{I})_{\text{free}}]^n}{K_{\text{Cu}}^n + [\text{Cu}(\text{I})_{\text{free}}]^n} \quad (3)$$

The maximum calculated Cu(I) occupancies of Csp1 were 11.3 and 11.7 equivalents for the BCS and BCA experiments at 101 and 1,210 μM respectively. Cu(I)-Csp1 samples with the maximum occupancy (for BCS from the titration and using samples prepared specifically for BCA) were separated from $[\text{Cu}(\text{L})_2]^{3-}$ and free L using a PD10 column in 20 mM HEPES pH 7.5 plus 200 mM NaCl. The protein and copper contents of the Cu(I)-Csp1-containing fractions were analysed with Bradford assays, and using BCS in the presence of >7 M urea. 11.6 ± 0.6 equivalents of Cu(I) per Csp1 were determined for samples from the experiment with 101 μM BCS to which 24.2–25.5 equivalents of Cu(I) were added (11.0 – 11.3 equivalents calculated using equation (1)), and 12.8 ± 0.7 for a Csp1 sample in

the presence of 1,200 μM BCA to which 18.4 equivalents of Cu(I) were added (11.3 – 11.6 equivalents calculated using equation (1)). In experiments with BCA, $[\text{Cu}(\text{I})_{\text{Csp1}}]$ appeared to decrease at higher Cu(I) concentrations, and this effect was much greater at higher BCA concentrations. For the data shown in Fig. 4b at 1,210 μM BCA, $[\text{Cu}(\text{I})_{\text{Csp1}}]$ decreased by <10% of the maximum value when 28.9 equivalents of Cu(I) were added (~20% upon addition of 49.5 equivalents of Cu(I) to a separate apo-Csp1 sample). In an experiment at 2,000 μM BCA, the maximum number of Cu(I) equivalents bound by Csp1 calculated using equation (1) was 9.33, achieved upon addition of 19.3 equivalents of Cu(I) and $[\text{Cu}(\text{I})_{\text{Csp1}}]$ appeared to decrease by almost 50% upon addition of 49.6 equivalents of Cu(I). The fit of the data up to 19.3 Cu(I) equivalents added, to the nonlinear Hill equation, gives $K_{\text{Cu}} = (6.3 \pm 0.1) \times 10^{-18} \text{ M}$ ($n = 4.1 \pm 0.2$), consistent with the values at lower BCA concentration (see Fig. 4c). A comparable apo-Csp1 sample in the presence of 2,000 μM BCA plus 59.9 equivalents of Cu(I) that gave 5.18 Cu(I) equivalents calculated using equation (1), was found to contain 11.6 ± 0.6 equivalents of Cu(I) after desalting on a PD10 column. We are currently unsure of the reason for the apparent decrease in $[\text{Cu}(\text{I})_{\text{Csp1}}]$, but it could be the result of the formation of Csp1-Cu(I)-BCA adducts. Such species may contribute to the absorbance at 562 nm, resulting in an apparent decrease in $[\text{Cu}(\text{I})_{\text{Csp1}}]$. However, this has little effect on the data at 1,210 μM BCA and the agreement in K_{Cu} for this experiment and that at 101 μM BCS (see Extended Data Fig. 6d) is very good (a fit of a repeat of the experiment with BCS to the nonlinear Hill equation gave $K_{\text{Cu}} = (1.3 \pm 0.1) \times 10^{-17} \text{ M}$ ($n = 2.4 \pm 0.2$)).

Analytical gel-filtration chromatography of apo- and Cu(I)-Csp1. Analytical gel-filtration chromatography of apo-Csp1 (~10–100 μM) and protein (~2–120 μM) plus ~12–14 equivalents of Cu(I) was performed on a Superdex 75 10/300 GL column equilibrated in 20 mM HEPES pH 7.5 plus 200 mM NaCl degassed and purged with nitrogen²⁴. Injection volumes ranged from 100 to 350 μl , the flow rate was 0.8 ml min⁻¹ and absorbance was monitored at 280 nm. Apparent molecular masses of 51 ± 3 ($n = 21$) and 50 ± 3 ($n = 18$) kDa for apo- and Cu(I)-Csp1 respectively were calculated from elution volumes by calibrating the column with a low molecular mass calibration kit (GE Healthcare)²⁴. The gel-filtration analysis of apo-Csp1 (~70–150 μM) plus ~22–26 equivalents of Cu(I) was performed three times and the eluted fractions were quantified for protein with Bradford assays, and for copper by AAS and Cu(I) using BCS in the presence of 7.6 M urea.

Circular dichroism spectroscopy. Far-UV circular dichroism spectra (180–250 nm) were recorded using a JASCO J-810 spectrometer^{36,41}. Apo-Csp1 and protein plus 14.0 equivalents of Cu(I) were analysed in 20 mM HEPES pH 7.5 plus 200 mM NaCl and in 100 mM potassium phosphate pH 8 (buffer exchanged using a PD10 column). Protein concentrations ranged from 7.94 to 39.7 μM (from 0.10 to 0.50 mg ml⁻¹). The pH-stability of apo-Csp1 in 20 mM buffer pH 5.5 (MES), 7.5 (HEPES) and 9.5 (CHES), all plus 200 mM NaCl, was monitored during 43 h incubation in the anaerobic chamber, with the final pH values of samples within 0.1 pH unit of the expected value (repeated at least twice). The α -helical content of apo- and Cu(I)-Csp1 was routinely found to be ~80% (repeated more than ten times). The unfolding of apo-Csp1 and protein plus 14.0 equivalents of Cu(I) was monitored after the addition of urea (final concentration >7 M) at pH 7.5.

Cu(I) exchange between Csp1 and mbtin. Cu(I)-Csp1 (~0.8–1.0 μM) loaded with ~13 equivalents of Cu(I) was added to apo-mbtin (~10–13 or ~20–27 μM). UV-vis spectra were acquired before the addition of Cu(I)-Csp1 and up to 6 h after mixing (repeated three times). Controls were performed by the addition of ~10–13 μM Cu(I) to apo-mbtin (~10–13 or ~20–27 μM). The possibility of copper transfer from Cu(I)-mbtin (typically ~2.2–3.0 μM) to apo-Csp1 (from 143 to 234 μM) was investigated for up to 20 h by UV-vis spectroscopy (repeated four times). All of the mbtin-containing samples were incubated anaerobically protected from light.

Crystallization, data collection, structure solution and refinement. Crystals of apo-Csp1 (~20 mg ml⁻¹, Bradford assay) in 20 mM HEPES pH 7.5 were obtained aerobically at 20 °C using the hanging drop method of vapour diffusion from 1 μl of protein mixed with 1 μl of 0.1 M bis-Tris pH 6.5 plus 25% w/v PEG 3350 (500 μl well solution) and were frozen in Paratone-N oil. Cu(I)-Csp1 (~9.5–11 mg ml⁻¹, Bradford assay) was prepared by adding ~12–14 equivalents of Cu(I) to apo-protein (~70–80 μM , quantified by DTNB) in 20 mM HEPES pH 7.5 plus 200 mM NaCl and subsequently concentrated by ultrafiltration (all performed anaerobically). Cu(I)-Csp1 prepared anaerobically was removed from the anaerobic chamber for <20 min (Cu(I)-Csp1 shows no sign of oxidation even after incubation in air for 42 h) to allow screens to be set up with a crystallization robot. The crystallization trays were transferred back into the anaerobic chamber via a port that was purged with nitrogen for 3 min and were in the chamber for <5 min before being sealed. Cu(I)-Csp1 crystallized at room temperature using the sitting drop method of vapour diffusion with 200 nl of protein plus 100 nl of 0.03 M MgCl₂, 0.03 M CaCl₂, 0.1 M Tris-Bicine pH 8.5, 12.5% v/v 2-methyl-2,4 pentanediol (racemic) plus 12.5% PEG 1000 and 12.5% PEG 3350 (80 μl well solution).

The Cu(I)-Csp1 crystal for the X-ray absorption near-edge spectrum shown in Extended Data Fig. 2c was obtained as above but using 600 nl of protein plus 300 nl of 0.025 M MgCl₂, 0.025 M CaCl₂, 0.1 M Tris-Bicine pH 8.5, 13.5% v/v 2-methyl-2,4-pentanediol (racemic) plus 13.5% PEG 1000 and 13.5% PEG 3350 (80 µl well solution). Crystals were frozen directly in the reservoir solution.

Diffraction data were collected at the Diamond Light Source, UK, on beamlines I02 (apo-Csp1, $\lambda = 0.9795$ Å) and I24 (Cu(I)-Csp1, $\lambda = 1.3777$ Å) at 100 K, processed and integrated with XDS and scaled using Aimless^{42,43}. For both data sets, space groups were determined using Pointless and later confirmed during refinement⁴⁴. The phase was solved by single-wavelength anomalous dispersion using copper, but was complicated by poorly resolved low-resolution reflections. The omission of data from 44.42 to 10.00 Å was required for successful heavy-atom location and the calculation of initial phases. Phasing, density modification and initial model building were performed using PHASER_EF⁴⁵ through the CCP4 interface⁴⁶, using SHELXD⁴⁷, PARROT⁴⁸ and BUCCANEER⁴⁹. The model of Cu(I)-Csp1 was used as the search model for molecular replacement in Molrep⁵⁰ to solve the apo-protein data set. The first 11 residues could not be modelled in both structures (His12 is close to the open end of an adjacent monomer in the Csp1 tetramer). Solvent molecules were added using COOT and checked manually. Simple solvent scaling was used for the apo-Csp1 model and Babinet solvent scaling was used for the Cu(I)-Csp1 model. All other computing used the CCP4 suite of programs⁴⁶. Five per cent of observations were randomly selected for the R_{free} set. The models were validated using MolProbity⁵¹ and data statistics and refinement details are reported in Extended Data Table 1. In a Ramachandran plot, 100% of residues are in most favoured regions for both models, and chain A of apo- and Cu(I)-Csp1 overlay with a root mean squared deviation of 0.42 Å. In the structure of Cu(I)-Csp1 (5AJF), the Cu(I) ions referred to herein as Cu1–Cu13 are numbered A1123–A1135 in chain A and the corresponding sites in chain B are numbered B1123–B1135.

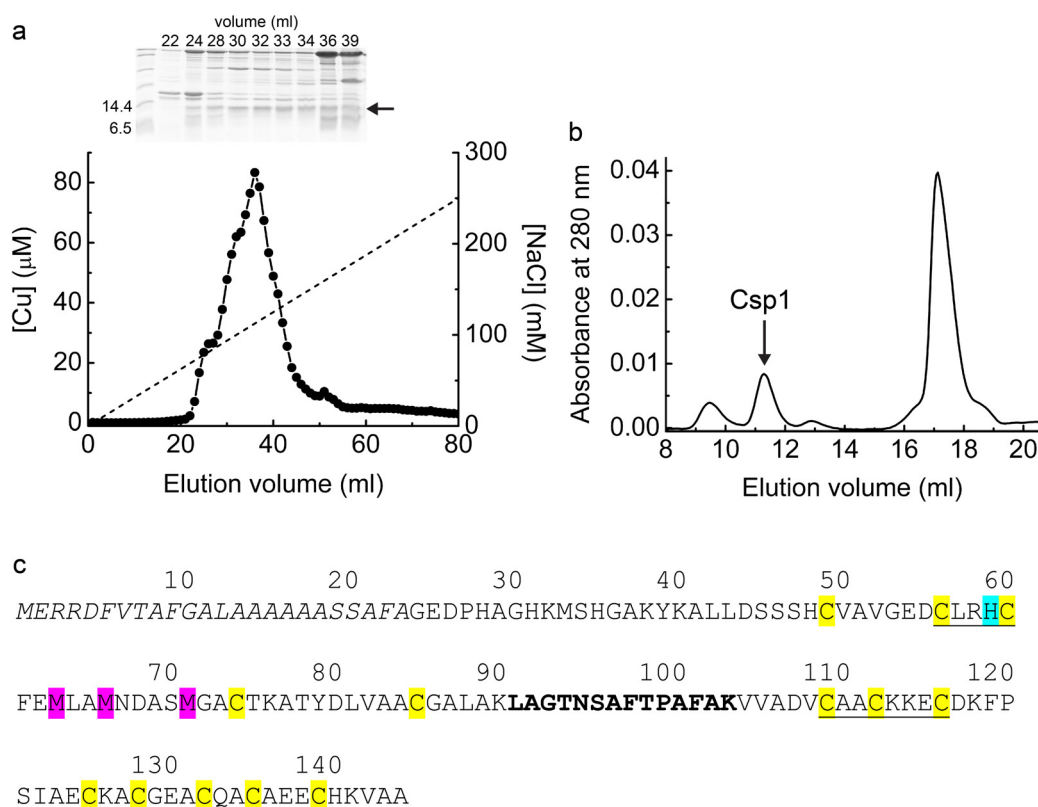
X-ray absorption near-edge spectroscopy. X-ray absorption near-edge spectroscopy was conducted on beamline I24 at the Diamond Light Source, UK, using a Vortex-EX detector (Hitachi). X-ray fluorescence was measured on a fresh Cu(I)-Csp1 crystal between 8,948 and 9,030 eV with an acquisition time of 3 s per data point and a constant step of 0.5 eV for the spectrum shown in Extended Data Fig. 2c (measurements were made on at least two other crystals giving very similar spectra).

Construction of strain $\Delta\text{csp1/csp2}$ of *M. trichosporium* OB3b. A double mutant, strain $\Delta\text{csp1/csp2}$, was constructed by sequential deletion of *csp1* followed by *csp2*, from *M. trichosporium* OB3b, using a previously described method⁵² with minor modifications. In each case, using genomic DNA from *M. trichosporium* OB3b as template, upstream and downstream regions (approximately 500 base pairs) of the target were amplified by PCR using primers (see Extended Data Table 2) 684AF/684AR and 684BF/684BR (for *csp1*) and 1592AF/1592AR and 1592BF/1592BR (for *csp2*). The resulting fragments were cloned into pK18mobsacB (ref. 52), which was then used to transform *E. coli* strain S17.1 (ref. 53). The constructs were introduced into *M. trichosporium* OB3b by conjugation as previously described⁵⁴ except that nalidixic acid was not required to remove *E. coli* contamination. Single crossover mutants were selected on NMS plates containing kanamycin (12.5 µg ml⁻¹). After cultivation in liquid medium without selection, double crossover mutants, with a deletion of the target gene, were selected by plating on NMS plates containing sucrose (7.5% w/v). Gene deletion was confirmed by PCR using primers (outside the cloned regions) 684TF/684TR2 (for *csp1*) and 1592TF/1592TR (for *csp2*) and sequencing.

sMMO activity of wild-type and mutant strains of *M. trichosporium* OB3b. The wild type and strain $\Delta\text{csp1/csp2}$ were grown in triplicate at 30 °C in NMS medium (50 ml), containing 6 µM copper, in 250 ml flasks supplied with 20% (v/v) methane and agitated at 150 r.p.m. At late exponential phase ($A_{540\text{ nm}} = 0.8$ – 1.0), cells were harvested by centrifugation at 5,000g for 15 min (22 °C), washed once and resuspended in NMS copper-free medium to a density of 1.5–1.6 ($A_{540\text{ nm}}$). Cell suspension (22 ml) was transferred to fresh 250 ml flasks and incubated with 20% v/v methane. Approximately every 3 h, samples (1 ml) were withdrawn and used to measure culture density ($A_{540\text{ nm}}$) and sMMO activity. After 12.75 and 19.25 h, flasks were flushed with air and re-supplied with methane. To estimate sMMO activity, cells (approximately 0.5 ml) were incubated with a few crushed crystals of naphthalene at 30 °C for 30 min before addition of 40 µl of tetrazotized *o*-dianisidine (10 mg ml⁻¹). Immediate development of an intense purple colour indicated sMMO activity⁵⁵. To quantify sMMO activity, 150 µl of cell suspension

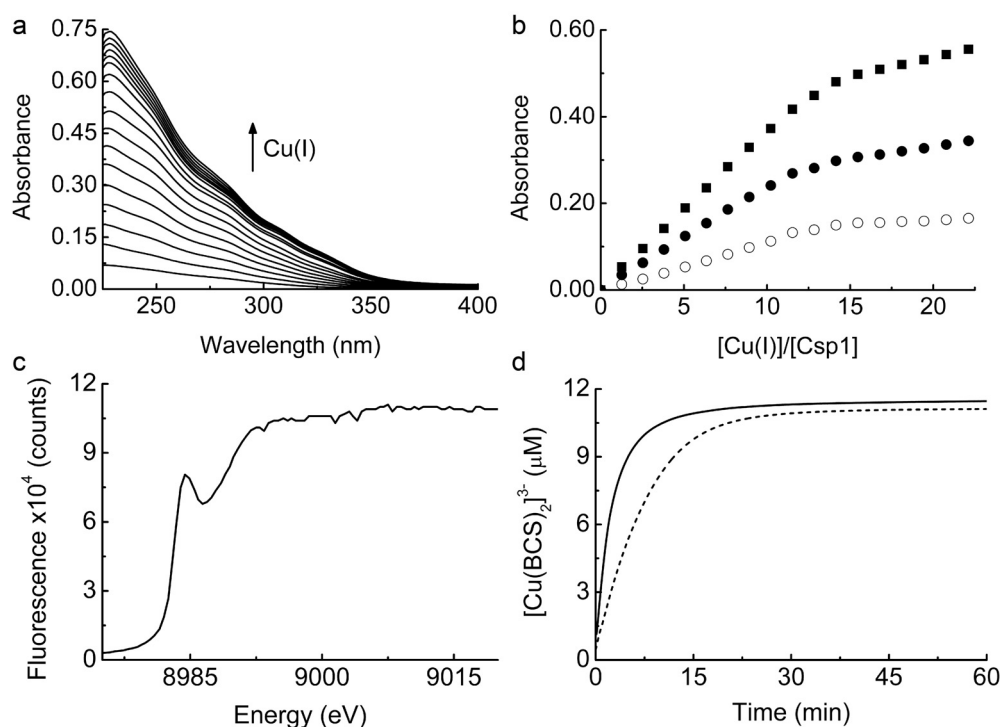
was centrifuged at 6,000g for 10 min (22 °C) and the cell pellet resuspended in 1 ml of 10 mM phosphate buffer pH 6.8 containing 10 mM formate. Crushed naphthalene crystals were added and the reaction initiated by addition of 25 µl of tetrazotized *o*-dianisidine (10 mg ml⁻¹). The mixture was shaken vigorously and the absorbance at 528 nm, corresponding to the formation of naphthol, was monitored for 15 min at 30 °C. Activities were normalized to cell density.

30. Tottey, S. *et al.* Protein-folding location can regulate manganese-binding versus copper- or zinc-binding. *Nature* **455**, 1138–1142 (2008).
31. Hellman, U., Wernstedt, C., Gonez, J. & Heldin, C. H. Improvement of an “In-Gel” digestion procedure for the micropreparation of internal protein fragments for amino acid sequencing. *Anal. Biochem.* **224**, 451–455 (1995).
32. Bendtsen, J. D., Nielsen, H., Widdick, D., Palmer, T. & Brunak, S. Prediction of twin-arginine signal peptides. *BMC Bioinform.* **6**, 167 (2005).
33. Riddles, P. W., Blakeley, R. L. & Zerner, B. Ellman’s reagent: 5,5'-dithiobis(2-nitrobenzoic acid)—a re-examination. *Anal. Biochem.* **94**, 75–81 (1979).
34. Riener, C. K., Kada, G. H. & Gruber, J. Quick measurement of protein sulphydryls with Ellman’s reagent and with 4,4'-dithiodipyridine. *Anal. Bioanal. Chem.* **373**, 266–276 (2002).
35. Allen, S., Badarau, A. & Dennison, C. Cu(I) affinities of the domain 1 and 3 sites in the human metallochaperone for Cu,Zn-superoxide dismutase. *Biochemistry* **51**, 1439–1448 (2012).
36. Badarau, A. & Dennison, C. Thermodynamics of copper and zinc distribution in the cyanobacterium *Synechocystis* PCC 6803. *Proc. Natl Acad. Sci. USA* **108**, 13007–13012 (2011).
37. Badarau, A. & Dennison, C. Copper trafficking mechanism of CXXC-containing domains: insight from the pH-dependence of their Cu(I) affinities. *J. Am. Chem. Soc.* **133**, 2983–2988 (2011).
38. Xiao, Z., Donnelly, P. S., Zimmerman, M. & Wedd, A. G. Transfer of copper between bis(thiosemicarbazone) ligands and intracellular copper-binding proteins. Insights into mechanisms of copper uptake and hypoxia selectivity. *Inorg. Chem.* **47**, 4338–4347 (2008).
39. Xiao, Z., Loughlin, F., George, G. N., Howlett, G. J. & Wedd, A. G. C-terminal domain of the membrane copper transporter Ctr1 from *Saccharomyces cerevisiae* binds four Cu(I) ions as a cuprous-thiolate polynuclear cluster: sub-femtomolar Cu(I) affinity of three proteins involved in copper trafficking. *J. Am. Chem. Soc.* **126**, 3081–3090 (2004).
40. Banci, L. *et al.* Affinity gradients drive copper to cellular destinations. *Nature* **465**, 645–648 (2010).
41. Allen, S., Badarau, A. & Dennison, C. The influence of protein folding on the copper affinities of trafficking and target sites. *Dalton Trans.* **42**, 3233–3239 (2013).
42. Kabsch, W. XDS. *Acta Crystallogr. D* **66**, 125–132 (2010).
43. Evans, P. R. & Murshudov, G. N. How good are my data and what is the resolution? *Acta Crystallogr. D* **69**, 1204–1214 (2013).
44. Evans, P. R. Scaling and assessment of data quality. *Acta Crystallogr. D* **62**, 72–82 (2006).
45. McCoy, A. J. *et al.* Phaser crystallographic software. *J. Appl. Cryst.* **40**, 658–674 (2007).
46. Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallogr. D* **67**, 235–242 (2011).
47. Sheldrick, G. M. Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. *Acta Crystallogr. D* **66**, 479–485 (2010).
48. Cowtan, K. Recent developments in classical density modification. *Acta Crystallogr. D* **66**, 470–478 (2010).
49. Cowtan, K. The Buccaneer software for automated model building. *Acta Crystallogr. D* **62**, 1002–1011 (2006).
50. Vagin, A. & Teplyakov, A. MOLREP: an automated program for molecular replacement. *J. Appl. Cryst.* **30**, 1022–1025 (1997).
51. Chen, V. B. *et al.* MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. D* **66**, 12–21 (2010).
52. Schäfer, A. *et al.* Small mobilizable multi-purpose cloning vectors derived from the *Escherichia coli* plasmids pK18 and pK19: selection of defined deletions in the chromosome of *Corynebacterium glutamicum*. *Gene* **145**, 69–73 (1994).
53. Simon, R., Priefer, U. & Pühler, A. A broad host range mobilization system for *in vivo* genetic engineering: transposon mutagenesis in Gram negative bacteria. *Nature Biotechnol.* **1**, 784–791 (1983).
54. Stafford, G. P., Scanlan, J., McDonald, I. R. & Murrell, J. C. *rpoN*, *mmoR* and *mmoG*, genes involved in regulating the expression of soluble methane monooxygenase in *Methylosinus trichosporium* OB3b. *Microbiology* **149**, 1771–1784 (2003).
55. Brusseau, G. A., Tsien, H.-C., Hanson, R. S. & Wackett, L. P. Optimization of trichloroethylene oxidation by methanotrophs and the use of a colorimetric assay to detect soluble methane monooxygenase activity. *Biodegradation* **1**, 19–29 (1990).
56. Xie, F., Sutherland, D. E. K., Stillman, N. J. & Ogawa, M. Y. Cu(I) binding properties of a designed metalloprotein. *J. Inorg. Biochem.* **104**, 261–267 (2010).
57. Notredame, C., Higgins, D. G. & Heringa, J. T-Coffee: a novel method for fast and accurate multiple sequence alignment. *J. Mol. Biol.* **302**, 205–217 (2000).



Extended Data Figure 1 | Purification of proteins from *M. trichosporium* OB3b and the amino-acid sequence of Csp1. **a**, The copper content of anion-exchange fractions (NaCl gradient shown as a dashed line) and the SDS-PAGE analysis of selected fractions (1 ml) from the purification of soluble extract from *M. trichosporium* OB3b cells. The band just below the 14.4 kDa marker, indicated with an arrow, is present. Fraction 32 was judged to have the lowest level of contaminating proteins and was further purified by gel-filtration chromatography on a Superdex 75 column (**b**). Csp1 is present in the peak that elutes at ~11 ml and contains considerable copper (see Fig. 1c). **c**, The amino-acid sequence of Csp1 showing the predicted Tat leader peptide (the first 24 residues of the pre-protein) in italics. The 13 Cys residues are highlighted in yellow, and His36 (cyan), Met40, Met43 and Met48 (magenta) are also

indicated (the numbering of these residues refers to the mature protein). The CXXXC and CXXC motifs are underlined. The region in bold corresponds to the single tryptic fragment identified on two separate occasions in MS analysis, representing 11% sequence coverage of the mature protein (Mascot search of peptide mass fingerprint, expectation value = 1.9×10^{-5}). The sequence of this fragment was confirmed by liquid chromatography/MS/MS (data not shown). This is the only tryptic peptide from the mature protein that would be anticipated to be readily detected by MS (owing to either small mass or presence of Cys residues in all other theoretical tryptic fragments) and is unique to this protein among all proteobacterial protein sequences in the NCBI database.



Extended Data Figure 2 | Cu(I) binding to Csp1. **a**, UV-vis difference spectra upon the addition of Cu(I) to apo-Csp1 (5.32 μM) showing the appearance of S(Cys) \rightarrow Cu(I) LMCT bands^{10,23,24}. **b**, Plots of absorbance at 250 nm (filled squares), 275 nm (filled circles) and 310 nm (open circles) against $[\text{Cu(I)}]/[\text{Csp1}]$ ratio taken from the spectra in **a**. The absorbance rises steeply until ~ 11 – 15 Cu(I) equivalents but continues to rise, particularly at lower wavelengths, making binding stoichiometry difficult to determine precisely with this approach. Systems that bind multiple Cu(I) ions in clusters such as those found in metallothioneins, typically give rise to luminescence at around 600 nm (refs 10, 13). However, limited luminescence is observed at

600 nm during the titration of Cu(I) into Csp1 (data not shown). **c**, X-ray absorption near-edge spectrum of a fresh crystal of Cu(I)-Csp1 at 100 K. **d**, Plots of $[\text{Cu(BCS)}_2]^{3-}$ formation against time after the addition of Cu(I)-Csp1 (0.93 μM) loaded with 11.8 equivalents of Cu(I) to 2,510 μM BCS either in the absence (dashed line) or presence (solid line) of 7.9 M urea. Cu(I) is removed faster in urea and is limited by the rate of Cu(I)-Csp1 unfolding (Extended Data Fig. 5i). The presence of urea has little effect on the end point for this reaction. Experiments in **a**, **b** and **d** were all performed in 20 mM HEPES pH 7.5 containing 200 mM NaCl.


```

Csp1      --GEDPHAGHKM-SHGAKYKALLDSSSHCVAVGEDCLRHC-----FEMLAMNDASMGACTKATYDLVAACGAL
Csp2      QTTQGLAPGAPVHHHPAKYHALMETSACVSTGNECLRHC-----FGMLSMNDTSMADCKASYDLVAACAAL
Csp3      MHVEAMISKHPQ-ARGQTDRSL---VQCVEMCFDCAQTCAACCADACLGEDKV--ADLRHCIRLNLDCAEICVAA
consensus  :      .      :      .  : : *      : * *      : * : *      :      :  : : * : . * . * *

```

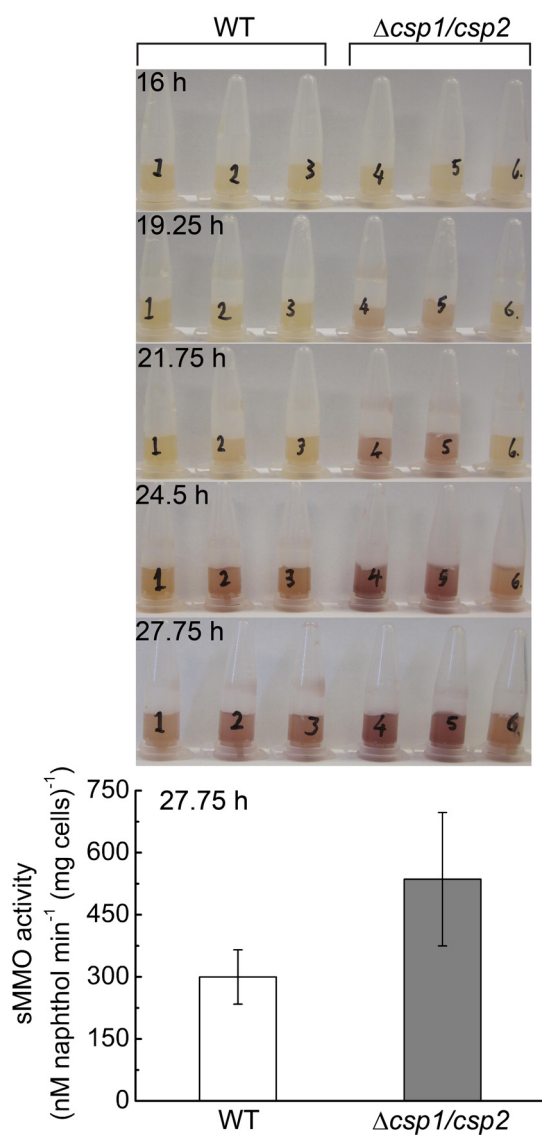
```

Csp1      AK----LAGTNSAFTPAFAKVVADVCAACCKKECDKFPS-IAECKACGEACQACAEECCHKVAA---
Csp2      ET----LSAVNSSATPALAKTVYDVCMACKKECDRFPQ-YSECKNCGDACKACCADECQRVSS---
Csp3      GSIASRAAGTEESILRTMLQTCAEMCRMCEEECDRRHAGNHEHCRICADVCKECETACRSATGLTH
consensus  .      : : : : :      : : . .      : : *      * : : * *      : : .      . * : * . : * : * : * : * : : .

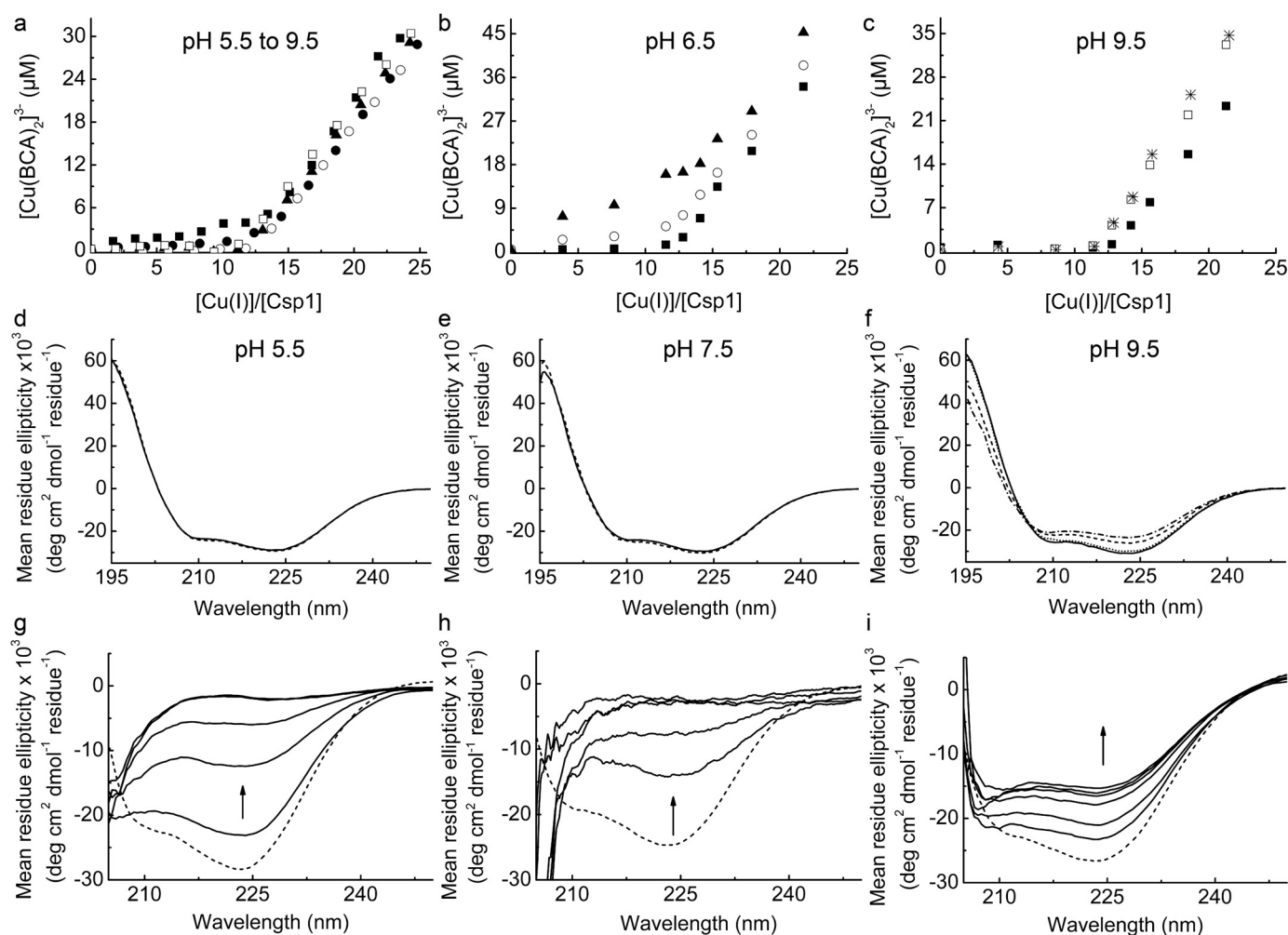
```

Extended Data Figure 3 | Sequence comparison of Csp1 homologues in *M. trichosporium* OB3b. The *M. trichosporium* OB3b genome possesses two genes that code for Csp1 homologues, Csp2 and Csp3, having 58 and 19% sequence identity to Csp1, respectively. The predicted Tat leader peptides of Csp1 (MERRDFVTAFGALAAAAASSAFA) and Csp2 (MERRQFVAAIGA AAAAASASRAFA) are omitted. The Cys residues (13 in Csp1 and Csp2 and 18 in Csp3) are highlighted in yellow with CXXXC and CXXC motifs underlined. A CXXXC motif in an α -helix allows both of the Cys residues to

coordinate the same Cu(I) ion (Fig. 3d, e), which is not the case for a CXXC motif. This is consistent with the observation that a synthetic peptide containing a CXXC motif binds a Cu_4S_4 cluster via a four-helix bundle made from four peptides, with coordination involving only one Cys per peptide^{28,56}. The alignment was produced using the T-coffee alignment tool⁵⁷. Asterisks indicate fully conserved sequence positions; the ':' and '.' symbols indicate strongly and weakly similar sequence positions respectively.

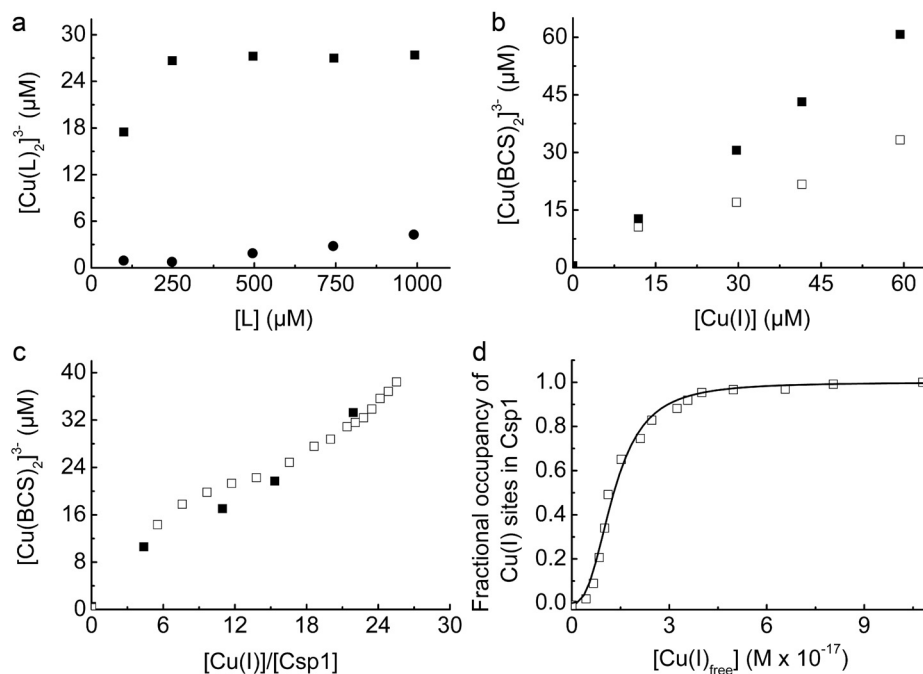


Extended Data Figure 4 | sMMO activity of wild-type *M. trichosporium* OB3b and the $\Delta csp1/csp2$ strain. Purple colour, indicating sMMO activity, is evident at 19.25 h in the $\Delta csp1/csp2$ strain (tubes 4–6), but not until 24.5 h in the wild type (WT, tubes 1–3) when using a qualitative assay. When quantified spectrophotometrically at 27.75 h, the average sMMO activity in the $\Delta csp1/csp2$ strain (grey) is 1.8-fold greater ($P = 0.04$, one-tailed t -test) than that of the wild type (WT, white), as shown in the bar chart (mean \pm s.d. of three replicates).



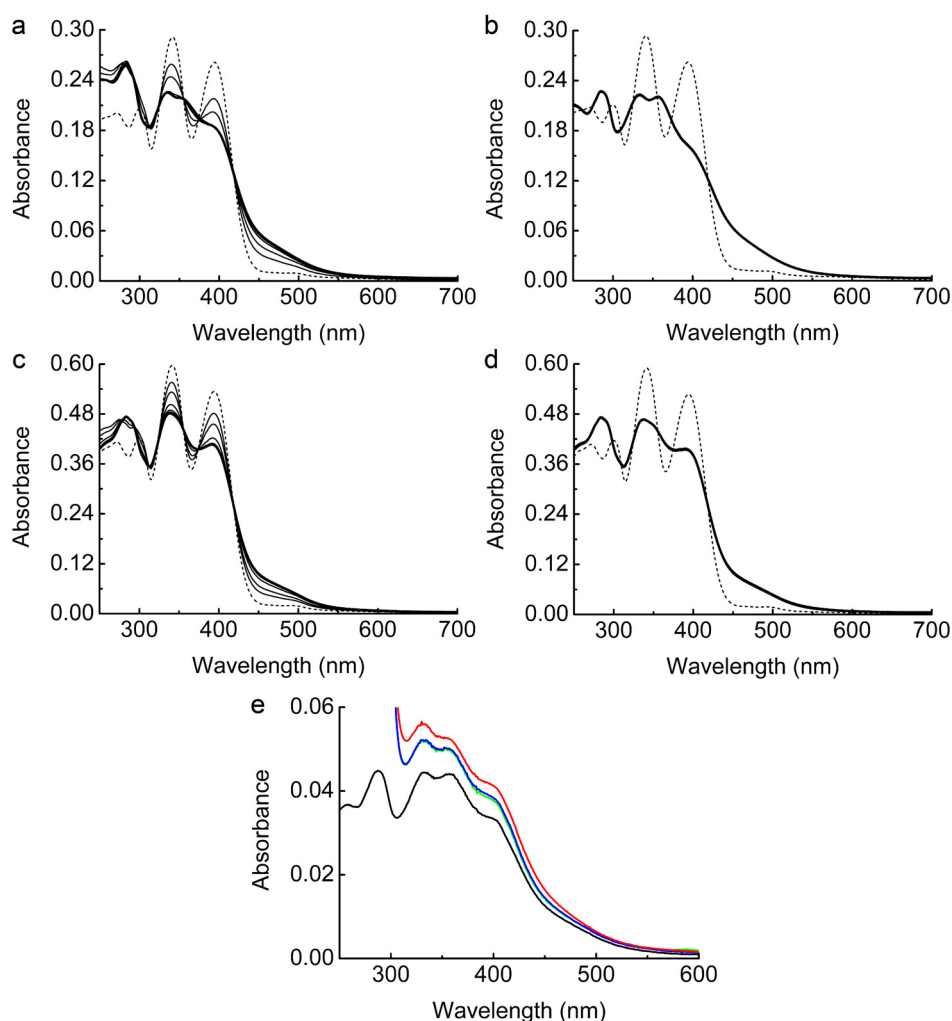
Extended Data Figure 5 | The dependence on pH of competition between Csp1 and BCA for Cu(I), and far-UV circular dichroism spectra showing pH stability and unfolding of Csp1 in urea. **a**, Plots of $[\text{Cu}(\text{BCA})_2]^{3-}$ concentration against $[\text{Cu(I)}]/[\text{Csp1}]$ ratio for the addition of Cu(I) to apo-Csp1 (2.38–2.56 μM) in the presence of 103 μM BCA in 20 mM buffer (see Methods) at pH 5.5 (filled squares), 6.5 (filled circles), 7.5 (filled triangles), 8.5 (open circles) and 9.5 (open squares) plus 200 mM NaCl. Equilibration is fast (< 20 min) at pH 6.5 and higher and the data shown are from titrations of Cu(I) into apo-Csp1. At pH 5.5 equilibration is slower and the data are for mixtures incubated for 21 h. Also shown are results for mixtures of Cu(I) with apo-Csp1 (3.31–3.67 μM) at pH 6.5 (b) and 9.5 (c) in the presence of 120 μM (filled squares), 300 μM (open circles), 450 μM (stars), 600 μM (filled triangles) and 900 μM (open squares) BCA, all after incubation for 15 h. At lower BCA concentrations, Csp1 is able to compete effectively for Cu(I) in the pH range 6.5–9.5, giving Cu(I) binding stoichiometries of 12–14 (see also Figs 3a and 4a). At pH 5.5, Csp1 competes less effectively with BCA for Cu(I), most probably because of the protonation of Cys ligands³⁷. This is consistent with greater competition by 600 μM BCA at pH 6.5 (b) compared with pH 7.5 (Fig. 4a). The stability of apo-Csp1 over the pH and time ranges used for experiments with BCA (and BCS) was determined using far-UV circular dichroism spectroscopy. The spectra of apo-Csp1 (solid lines) at pH (d) 5.5 (34.1 μM , 0.43 mg ml^{-1}), (e) 7.5 (36.5 μM , 0.46 mg ml^{-1}) and (f) 9.5 (32.6 μM ,

0.41 mg ml^{-1}) are compared with those for samples incubated for 43 h (dashed lines), and for 3 h (dotted line) and 17 h (dashed/dotted lines) at pH 9.5. At pH 9.5 and in the presence of higher BCA concentrations (c), Csp1 binds approximately one less equivalent of Cu(I), which must be because of changes in structure that are observed at this pH value (no change after 3 h but there is a decrease of ~ 15 –20% α -helical content at longer incubation times, see f). However, the remaining sites bind Cu(I) more tightly (c) than at pH 7.5 (Fig. 4a) because of deprotonation of the Cys ligands³⁷. **g**, Far-UV circular dichroism spectra of apo-Csp1 (19.9 μM , 0.25 mg ml^{-1}) in 20 mM HEPES pH 7.5 containing 200 mM NaCl at 0, 30, 60, 120 and 240 min (solid lines) after the addition of urea (7 M) compared with the spectrum for apo-Csp1 in the same buffer but with no urea (dashed line). **h**, Far-UV circular dichroism spectra of apo-Csp1 (7.94 μM , 0.10 mg ml^{-1}) as in g except that spectra were acquired at 0, 15, 30, 45 and 60 min (solid lines) after addition of urea (7 M); unfolding is significantly faster at lower protein concentrations and is consistent with the reaction with DTNB in urea being complete in 20 min at Csp1 concentrations $< 4 \mu\text{M}$. **i**, Far-UV circular dichroism spectra of Csp1 incubated with 14.0 equivalents of Cu(I) (19.9 μM , 0.25 mg ml^{-1}) as in g but at 0, 60, 240, 360 and 480 min and 24 h (solid lines) after addition of urea (7 M) compared with the spectrum for Cu(I)-Csp1 in buffer with no urea (dashed line). The arrow in g to i indicates how the spectrum changes with time.



Extended Data Figure 6 | Competition for Cu(I) between Csp1 and chromophoric ligands and the determination of the apparent average Cu(I) dissociation constant for Csp1 using BCS. **a**, Plots of $[\text{Cu}(\text{L})_2]^{3-}$ concentration against $[\text{L}]$ (BCA or BCS) after the incubation of Cu(I)-Csp1 (2.59 μM) loaded with 10.4 equivalents of Cu(I) with different concentrations of BCA (filled circles) and BCS (filled squares) for 17 h. **b**, Plots of $[\text{Cu}(\text{BCS})_2]^{3-}$ concentration against $[\text{Cu}(\text{I})]$ for apo-Csp1 (2.71 μM) in the presence of 99.2 μM (open squares) and 248 μM (filled squares) BCS incubated with increasing concentrations of Cu(I) (0, 4.38, 11.0, 15.3 and 21.9 equivalents; data shown after 17 h incubation). BCS competes much more effectively with Csp1 for Cu(I) than BCA, and $[\text{Cu}(\text{BCS})_2]^{3-}$ is stoichiometrically formed at 248 μM BCS. **c**, Plot of $[\text{Cu}(\text{BCS})_2]^{3-}$ concentration against the

$[\text{Cu}(\text{I})]/[\text{Csp1}]$ ratio for mixtures of Cu(I) plus apo-Csp1 (2.70 μM) in the presence of 101 μM BCS (open squares) for 19 h. For comparison, the data from **b** (2.71 μM Csp1 in the presence of 99.2 μM BCS for 17 h) are also shown (filled squares). The data in **a–c** were all acquired in 20 mM HEPES plus 200 mM NaCl at pH 7.5. **d**, Fractional occupancy of Cu(I)-binding sites in Csp1 (maximum value is 11.3 equivalents in this experiment) at different concentrations of free Cu(I) for the experiment shown in **c**. The solid line shows the fit of the data to the nonlinear Hill equation giving $K_{\text{Cu}} = (1.3 \pm 0.1) \times 10^{-17} \text{ M}$ ($n = 2.7 \pm 0.2$). Hill coefficients larger than 1 indicate positive cooperativity for Cu(I) binding by Csp1. Confirmation, and the cause, of this effect will be the subject of further studies.



Extended Data Figure 7 | Cu(I) exchange between Csp1 and mbttin. UV-vis spectra of apo-mbttin (dashed lines) and at various times up to 360 min (thick lines) after the addition of either Cu(I)-Csp1 or Cu(I). Cu(I)-Csp1 (1.02 μM) loaded with 13.0 equivalents of Cu(I) was added to either 13.4 μM (a) or 27.4 μM (c) apo-mbttin. Cu(I) alone (13.3 μM) was added to 13.4 μM (b) or 27.1 μM (d) apo-mbttin. Plots of absorbance at 394 nm against time for a–d are shown in Fig. 4d. Mbttin from *M. trichosporium* OB3b has a Cu(I) affinity of $(6\text{--}7) \times 10^{20} \text{ M}^{-1}$ at pH 7.5 (determined¹⁷ using a $\log\beta_2$ value of 19.8 for $[\text{Cu}(\text{BCS})_2]^{3-}$, but is an order of magnitude tighter if the more recent $\log\beta_2$ value of 20.8 (ref. 25) is used) and stoichiometrically removes Cu(I) from Csp1

within 1 h. e, UV-vis spectra of Cu(I)-mbttin (2.71 μM , black line) immediately after mixing with apo-Csp1 (234 μM , green line) and after incubation under anaerobic conditions for 1 h (blue line) and 20 h (red line). Small increases in absorbance are observed because of the absorbance of apo-Csp1 at these wavelengths and precipitation. The latter was more of a problem at longer incubation times and the sample at 20 h required filtering before running the spectrum shown. The small changes observed are not consistent with the formation of apo-mbttin¹⁷. All experiments were performed in 20 mM HEPES pH 7.5 plus 200 mM NaCl.

<i>N. gonorrhoeae</i>	MNRRQF--LGSAAAVSLASAASFARAHGHA-----D----YHHHHMQPAAASAYTAVRQTAAHCLDAGQVCL-THCLSLLTQGDTSMSDCAVAVRQM
OB3b Csp1	MERRDF--VTAFGALAAAAAASS--AFAGE-----D----PHAGHKMSH--GAKYKALLDSSSHCVAVGEDCL-RHCFEMLAMNDASMGACTKATYDL
OB3b Csp2	MERRQF--VAAIGAAAAAASASR--AFAQT-----TQGLAPGAPVH-HH--PAKYHALMETSAKCVSTGNECL-RHCFGMLSMNDTSMACTKASYDL
<i>P. aeruginosa</i>	MTRAINDPGNED-----PGSLLETADALLGGAAVQAPE-----ERCLASQDCIRACERYL-ALCISS-----REQRQHAGDC
OB3b Csp3	MHVEAM--ISKH-----PQARGQT-----DRSL-----VQVEMCFDCAQTCAACA-DACLGEDKV--ADLRHCIRLNLDC
<i>S. coelicolor</i>	MPTTVNDLLRTY-----PADLGGVD-----REAM-----ARCIEECLRCAQACTACA-DACLEPTV--ADLTKCIRTDMDC
<i>N. multiformis</i>	MFLYTE-----T-----DQNL-----QACIDACNHCYRTCLRMAMNHCLEAGGK--HVEADHLRLMMNC
<i>R. leguminosarum</i>	MHHMS-----TEM-----KACIDNCLACYSECLSMAMGHCLLELGE--HTKPPHFKLMMAC
<i>R. metallidurans</i>	MIRPTV--QE-----N-----FSRY-----ADCIACNAAAAACLKCA-AALEEPDT--RKMTRCIALDMDC
<i>S. enterica</i>	MQQ-E-----H-----RECIEQCYECAAACDICA-SSCLREDNV--EMMKHCQLDMQC
<i>B. subtilis</i>	MEQY-----S-----EACIEACIDCMKACNHC-F-TKLEESVQ--HHLSGCIRLDREC
<i>L. pneumophila</i>	MTHQQ-----Y-----DMCIKACQACLLECEHCA-NA--LHEEDC--NDLARCISLDRDC
consensus	* * * * *
<i>N. gonorrhoeae</i>	LALCGAVHDLAAQN----SPLTRDAKVCLEACKQCAKCKEHSAHHAECCKAYESCLDCKEKLAA---
OB3b Csp1	VAAAGALAKLAGTN----SAFTPAPAKVVADVCAACKKECDKFPS-IAECKAGGEACQACAEECHKVAA---
OB3b Csp2	VAAAGAALETLSAVN----SSATPALAKTVYDVCMACKKECDRFPPQ-YSECKNGGDACKACADECQVRSS---
<i>P. aeruginosa</i>	ADLCRLAALLERR---SPWAPACELAARYALACAERC DGDEP---LERECAGACRRFVEACRPLLPAA--
OB3b Csp3	AEICVAAGSIASRAAGTEESILRTMLQTCAEMCRMCEECRRHAGNHEHCRI CADVCKECETACRSATGLTH
<i>S. coelicolor</i>	ADVCTATAAVLSRHTGYDANVTRAVLQACATVCAACGDECARHAGMHEHCRCVCAEACRSCEQACQELLAGLG
<i>N. multiformis</i>	AEICQTSLNFMLSG---SRFSPKVCVGC AEICDACA KSC EQLDG---MEECVQTCRCQAEHCRKMAA---
<i>R. leguminosarum</i>	AEICRTSAHFMLIG---SEHHKHVCRECAEICGQCAEDCERVG---DMQSCVDACRRCADSCRKMAA---
<i>R. metallidurans</i>	AGIANLAASYMLRN---SEFAPLVCEDECAEVCKWCKECCERYDH--WHCQECAKACAACMEMCLKMTA---
<i>S. enterica</i>	AAICRLAAQFMALE---SEYSQKLCRLCADICKACAEECARHDH--DHCQNCARACSCQACADACLKMAA---
<i>B. subtilis</i>	ADICALAVKAMQTD---SPFMKEICALCADICEACGTECGKHDH--DHCQACAKACFTCAEQCRSMAA---
<i>L. pneumophila</i>	AAICALAIEMMARN---SPFAKEICALCAKICRACGDECSKHQH-MEHCQRCAKACYQCAEACEKM-A---
consensus	* * * * *

Extended Data Figure 8 | Sequence comparison of Csp homologues from diverse bacteria. Homology searches show that Csp homologues are encoded in the genomes of diverse bacteria. Multiple sequence alignment of the three *M. trichosporium* OB3b proteins (OB3b Csp1, OB3b Csp2 and OB3b Csp3) with a selection of these proteins, including one member (from *Neisseria gonorrhoeae*) that also possesses a putative Tat signal sequence (underlined), shows that the Cys residues (highlighted in yellow) are highly conserved. The alignment was produced using the T-coffee alignment tool⁵⁷. Asterisks indicate fully conserved sequence positions; the ‘.’ and ‘:’ symbols indicate strongly and weakly similar sequence positions respectively. *N. gonorrhoeae* sequence: open reading frame (ORF) NGAG_01502, UniProt accession

C1I025; *P. aeruginosa* sequence: ORF PA96_2930, UniProt accession X5E748 (PDB accession number 3KAW); *Streptomyces coelicolor* sequence: ORF SCO3281, UniProt accession Q9X8F4; *N. multiformis* sequence: ORF NmuI_A1745, UniProt accession Q2Y879 (PDB accession number 3LMF); *Rhizobium leguminosarum* sequence: ORF RLEG_20420, UniProt accession W0IHZ3; *Ralstonia metallidurans* sequence: ORF Rmet_5753, UniProt accession Q1LB64; *Salmonella enterica* sv. Typhimurium sequence: ORF STM14_1521, UniProt accession D0ZVJ6; *Bacillus subtilis* sequence: ORF BSU10600, UniProt accession O07571; *Legionella pneumophila* sequence: ORF LPE509_p00081, UniProt accession M4SK87.

Extended Data Table 1 | Data collection and refinement statistics

	Apo-Csp1	Cu(I)-Csp1
Data collection		
Space group	P2 ₁	P2
Cell dimensions		
<i>a</i> , <i>b</i> , <i>c</i> (Å)	40.9, 105.9, 48.7	44.4, 41.4, 53.1
α , β , γ (°)	90.0, 112.5, 90.0	90.0, 92.6, 90.0
Resolution (Å)	44.95-1.50 (1.53-1.50)*	53.06-1.90 (1.95-1.90)
<i>R</i> _{merge} (%)	7.0 (50.5)	8.7 (43.3)
<i>I</i> / σ <i>I</i>	10.9 (2.6)	5.6 (2.1)
Completeness (%)	99.7 (99.8)	99.1 (97.1)
Redundancy	3.7 (3.7)	2.8 (2.4)
Refinement		
Resolution (Å)	1.50	1.90
No. reflections	60896 (3056)	15212 (990)
<i>R</i> _{work} / <i>R</i> _{free}	12.2/17.9	19.8/23.2
No. atoms		
Protein	3209	1575
Ligand/ion	0	28
Water	406	116
B-factors		
Protein	16.2	40.2
Ligand/ion		41.4
Water	27.0	47.7
R.m.s deviations		
Bond lengths (Å)	0.020	0.016
Bond angles (°)	1.8	1.6

*Highest resolution shell is shown in parenthesis.

Extended Data Table 2 | Primers used for cloning Csp1 and making the $\Delta csp1/csp2$ *M. trichosporium* OB3b strain

Primer	Sequence (5' to 3')*
Csp1_F	GCGCATATGGGAGAGGATCCTCATGC
Csp1_R	GCGCCATGGTCAGGCGGCGACCTTATGGC
684AF	ATATCCCGGGTAAGGGTGAAG ACCGCCATCAG
684AR	GATCGTCGACACGACGGACGCAACCTAAAC
684BF	GATCGTCGACTAAGGTCGCCGCCTGAGTTC
684BR	GATCAAGCTTCGCGCTCGCGTCCGTATTC
1592AF	CATCAAGCTTCGGTGCGCGACATCATCCTC
1592AR	CATCCTGCAGTGGTCGTTCTCTCGTGTTTC
1592BF	TAATGGATCCCAGCGCGTGTGAGCTGAAC
1592BR	ATTAGAATTTCGCGGAGCCCGCGTGGAAG
684TF	CACATGCAGGCGGTAGATCG
684TR2	CGACCAGCAGGATCATCAG
1592TF	ACCCTTCTCACGCAATCCC
1592TR	ACGTTGATCGGCCTCACTC

* Introduced restriction sites are underlined when relevant.

CORRIGENDUM

doi:10.1038/nature14554

Corrigendum: A diverse range of gene products are effectors of the type I interferon antiviral response

John W. Schoggins, Sam J. Wilson, Maryline Panis, Mary Y. Murphy, Christopher T. Jones, Paul Bieniasz & Charles M. Rice

Nature **472**, 481–485 (2011); doi:10.1038/nature09907

We have recently discovered that the WNV-GFP stock used in the data set (Fig. 2 and Supplementary Table 8 of the original Letter) for West Nile virus (WNV) in this Letter was actually Venezuelan equine encephalitis virus (VEEV-GFP). The error has been tracked to a technical mistake made during the virus production process. This error affected the panels labelled WNV in Fig. 2, Supplementary Fig. 9, and Supplementary Table 8 of the original Letter; the Supplementary Information to this Corrigendum contains the corrected WNV data sets. This error also affected figure labels in our later work¹ (see the Corrigendum to ref. 1).

1. Schoggins, J. W. *et al.* Pan-viral specificity of IFN-induced genes reveals new roles for cGAS in innate immunity. *Nature* **505**, 691–695 (2014); corrigendum *Nature* <http://dx.doi.org/10.1038/nature14555> (2015).

Supplementary Information is available in the online version of the paper.

CORRIGENDUM

doi:10.1038/nature14555

Corrigendum: Pan-viral specificity of IFN-induced genes reveals new roles for cGAS in innate immunity

John W. Schoggins, Donna A. MacDuff, Naoko Imanaka, Maria D. Gainey, Bimmi Shrestha, Jennifer L. Eitson, Katrina B. Mar, R. Blake Richardson, Alexander V. Ratushny, Vladimir Litvak, Rea Dabelic, Balaji Manicassamy, John D. Aitchison, Alan Aderem, Richard M. Elliott, Adolfo García-Sastre, Vincent Racaniello, Eric J. Snijder, Wayne M. Yokoyama, Michael S. Diamond, Herbert W. Virgin & Charles M. Rice

Nature **505**, 691–695 (2014); doi:10.1038/nature12862

In this Letter, we carried out bioinformatic analyses on interferon-stimulated gene screening data sets for multiple viruses, including a data set for West Nile virus (WNV) (Supplementary Table 8 in ref. 1). We recently discovered that the WNV-GFP stock used in our 2011 study¹ was actually Venezuelan equine encephalitis virus (VEEV-GFP). The error has been tracked to a technical mistake made during the virus production process. Several data sets in this Letter are therefore mislabelled. In Fig. 3a and in all panels of Extended Data Fig. 2a, ‘WNV’ should be ‘VEEV’. The original figure legends remain valid, as do all the other figures in this Letter. One conclusion of the Letter highlighted differences in interferon-stimulated gene specificity between positive-sense and negative-sense RNA viruses. Since VEEV and WNV are both positive-sense, the stated conclusions remain unchanged; all other results and conclusions are also unchanged.

1. Schoggins, J. W. *et al.* A diverse range of gene products are effectors of the type I interferon antiviral response. *Nature* **472**, 481–485 (2011); corrigendum *Nature* <http://dx.doi.org/10.1038/nature14554> (2015).

CORRIGENDUM

doi:10.1038/nature14608

Corrigendum: Greenland supraglacial lake drainages triggered by hydrologically induced basal slip

Laura A. Stevens, Mark D. Behn, Jeffrey J. McGuire,
Sarah B. Das, Ian Joughin, Thomas Herring,
David E. Shean & Matt A. King

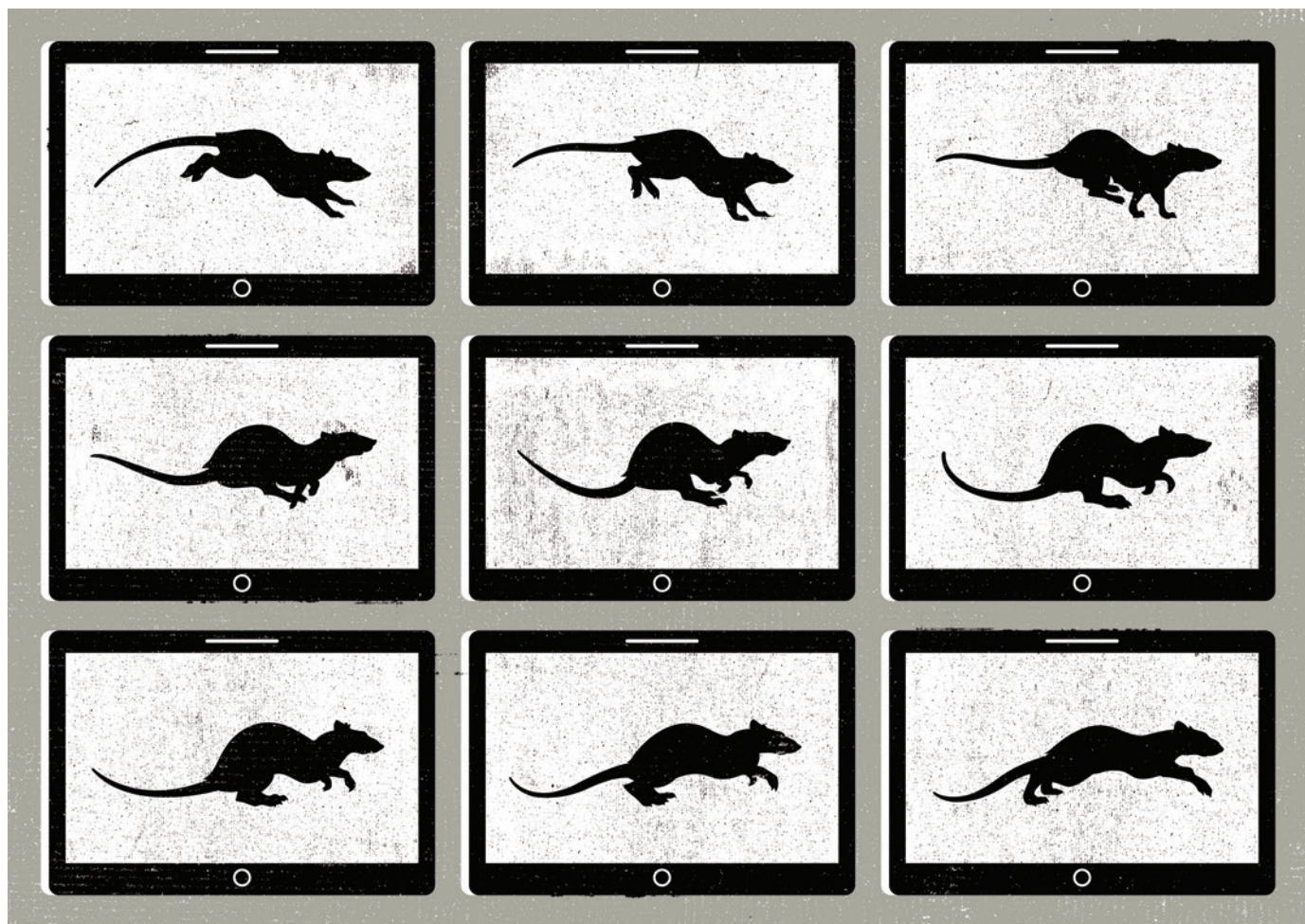
Nature **522**, 73–76 (2015); doi:10.1038/nature14480

In Fig. 2d and f of this Letter, the labels ' 10^8 N m' should read ' 10^{17} N m'.
(Basal moment is correctly reported in Extended Data Table 1.)
Figure 2 has been corrected online.

TOOLBOX SEE HOW THEY RUN

Software tools that track how animals move are helping researchers to do everything from diagnosing neurological conditions to illuminating evolution.

ILLUSTRATION BY THE PROJECT TWINS



BY BOER DENG

Palaeontologist Stephen Gatesy wants to bring extinct creatures to life — virtually speaking. When he pores over the fossilized skeletons of dinosaurs and other long-dead beasts, he tries to imagine how they walked, ran or flew, and how those movements evolved into the gaits of their modern descendants. “I’m a very visual guy,” he says.

But fossils are lifeless and static, and can only tell Gatesy so much. So instead, he relies

on XROMM, a software package that he developed with his colleagues at Brown University in Providence, Rhode Island. XROMM (X-ray Reconstruction of Moving Morphology) borrows from the technology of motion capture, in which multiple cameras film a moving

➔ **NATURE.COM**
For more on scientific software, apps and online tools, visit: nature.com/toolbox

object from different angles, and markers on the object are rendered into 3D by a computer program. The difference is that XROMM uses not

cameras, but X-ray machines that make videos of bones and joints moving inside live creatures such as pigs, ducks and fish. Understanding how the movements relate to the animals’ bone structure can help palaeontologists to determine what movements would have been possible for fossilized creatures. “It’s a completely different approach” to studying evolution, says Gatesy.

XROMM, released to the public in 2008 as an open-source package, is one of a number of software tools that are expanding what researchers know about how animals and ►

► humans walk, crawl and, in some cases, fly (see ‘Movement from inside and out’). That has given the centuries-old science of animal motion relevance to a wide range of fields, from studying biodiversity to designing leg braces, prostheses and other assistive medical devices. “We’re in an intense period of using camera-based and computer-based approaches to expand the questions we can ask about motion,” says Michael Dickinson, a neuroscientist at the California Institute of Technology in Pasadena.

To use and develop effective software, however, scientists must learn how to adapt broad, open-source tools to their own needs — and when to build their own.

A VISUAL HISTORY

The boom in motion-tracking tools has come about in part because of improvements in what researchers can see and measure. The first studies of animal and human motion, dating back to Aristotle, relied on naked-eye observations, anatomy and detailed pictures drawn by hand. In the nineteenth century, the science of biomechanics was boosted by photography — perhaps most famously in a series of images of a galloping horse taken by British photographer Eadweard Muybridge, and published in his *Animal Locomotion* collection in 1887.

Higher-speed cameras eventually improved what could be captured. But movement studies still needed a person to look through the results frame by frame, laboriously tracing the arc of each step, arm swing or wing flap to extract information about angles and forces. Much of that tedium can now be relieved by computers or other measuring tools. But such tools are often expensive, and even today, many researchers do without them. Gatesy recalls a graduate student’s surprise at the low-tech approach that was used to study gait in rodents a few years ago: “It wasn’t uncommon just to dip their feet into some ink, have them leave some tracks and take measurements from those,” he says.

Lately, however, scientists have been coming

up with methods that are much more sophisticated without being too expensive. In July, developmental biologists Richard Mann and César Mendes at Columbia University in New York City and their colleagues published a paper on MouseWalker: a system they have built to automatically analyse changes in a mouse’s gait (C. S. Mendes *et al. BMC Biol.* **13**, 50; 2015). It involves an inexpensive set-up in which a mouse walks on a transparent surface over a high-speed camera that records the animal’s footfalls. An analytical technology called machine vision allows the MouseWalker software to discern details such as the position of each step relative to the mouse’s body.

Mendes says that this information can be used to detect when something goes wrong with gait, as can happen with the onset of neurological illnesses such as Parkinson’s disease. MouseWalker was adapted from FlyWalker, a system that Mendes and his team helped to develop to let neuroscientists track how fruit flies walk after their neurons have been manipulated. Both MouseWalker and FlyWalker are open source: the authors hope that making the software available for free will help to attract users who can add parameters that they had not thought of.

BUILDING USER COMMUNITIES

The desire to share tools is common to many developers, so motion-tracking software is finding applications in a number of fields — sometimes in unexpected ways. “One of the things we hope for is that people will use what we develop and go in a new direction with it,” says Jen Hicks, an engineer at Stanford University in California, who helps to manage OpenSim — an open-source software package that allows users to model joints, muscles and how they move. OpenSim has more than 20,000 users, and part of Hicks’s job is to organize workshops and tutorials to guide this growing community.

The OpenSim community serves as an exemplar of what newer programs such as XROMM or MouseWalker could become. The

software models musculoskeletal systems, and researchers have used it to simulate everything from the potential outcomes of surgery to the muscular forces of goats. Since the first version of OpenSim was released in 2007, the package has gone through dozens of upgrades that have added features and improved the algorithms used for calculations. It has been downloaded more than 100,000 times. “It’s amazing how much the community has grown,” says mechanical engineer Katherine Steele of the University of Washington in Seattle, who first began using OpenSim while studying cystic fibrosis as a graduate student at Stanford.

Serving an ever-larger crowd requires careful planning to make the program accessible, says Hicks. Through grants from the US National Institutes of Health, she and her colleagues keep manuals up to date with the new releases. Ensuring that the software can be tailored to a researcher’s particular needs has helped new users to embrace it, she says.

THE LIMITS OF BROAD PLATFORMS

XROMM’s developers are in the middle of building up the infrastructure to make the software accessible to a wider community, for instance setting up a site to host the newest open-source version, XMA Lab, which became available in December 2014. The team has tried to make the latest versions of the software easier for new users. For example, says Elizabeth Brainerd, a colleague of Gatesy at Brown, “There used to be about 20 pieces of information you had to keep track of,” including items such as calibration measurements. “But now it’s all integrated.”

It is important not to make things too easy, says Steele: if the software does too much of the work, there is a risk that the researchers will misunderstand the data that it spits out. However, as an open-source program develops, understanding its architecture can get very complicated. “Sometimes the software can get so big that it becomes black-box-ish. Then it might be better for you to build your own,” she says.

Dickinson agrees, and says that sometimes, modifying open-source tools is not enough. “As science is becoming more quantitative, we’re all working on finer slices of the pie,” he says. “If you only got to use a microscope that someone else built, so to speak, you won’t be able to get as far.”

Regardless of what tools are available, researchers intend to keep expanding the applications of motion tracking. Hicks anticipates seeing more people using the tools to explore neural control and robotics designs. And she expects the software to keep improving. “We’re finding ways to learn from even messier motion data, like from accelerometers in your phone,” she says. “Bringing together more machine learning and biomechanics — that will be the next step.” ■

Boer Deng is a journalist in Washington DC.

RESOURCES

Movement from inside and out

Scientists monitoring animal motion use a variety of programs to automate the process.

- MouseWalker (go.nature.com/hugtga) is a gait-tracking system that can help researchers studying the connection between neuroscience and movement. The open-source software package was released in July. Previously, the developers collaborated on FlyWalker (go.nature.com/jfwgfo), a tool for measuring fly walking.
- X-ray Reconstruction of Moving Morphology, or XROMM (go.nature.com/58bn2s), helps researchers to visualize

animals’ bones and joints as 3D moving skeletons. XMA Lab (go.nature.com/wz1msi), the latest version of the XROMM software, was released in December 2014.

- OpenSim (go.nature.com/1ulnpq) is an open-source program that allows researchers to model muscles, bones and the forces that act on them. Some researchers have used it to simulate the outcomes of surgery or to test hypotheses about movement pathologies such as those that affect people with cystic fibrosis or Parkinson’s disease.

CAREERS

TEACHING COUNTS An ecosystem modeller takes an unusual track to tenure **p.149**

ADVICE Bring on the questions about life, work and science go.nature.com/monvth

NATUREJOBS For the latest career listings and advice www.naturejobs.com



PATRICK GEORGE/GETTY

INTERNSHIPS

Mind wide open

An innovative US National Institutes of Health programme aims to expose junior scientists to different career paths.

BY PAUL SMAGLIK

Internships are not just for undergraduates any more. With fewer than one-third of US life-science PhD graduates destined for tenure-track positions, most graduate students and postdoctoral researchers need to prepare themselves for life outside the academic laboratory.

In 2013, the US National Institutes of Health (NIH) launched the Broadening Experience in Scientific Training (BEST) programme, after being faced with a number of reports

suggesting that the US graduate-education system trains scientists for faculty positions that do not actually exist (see go.nature.com/buk6km). The initiative began as a five-year pilot to offer biomedical graduate students and postdocs supplemental skills to help to prepare them for non-tenure-track career options.

Since its launch, BEST has offered enhanced career training, including internships, to about 10,000 graduate students and 600 postdocs. The 17 universities that received NIH grants to participate in the pilot programme act as

a collective laboratory, exploring different approaches to redefine graduate training and craft internships for highly trained young scientists. The approaches may include partnerships with industry and other sectors; participant institutions are sharing their experiences to determine best practices. So far, BEST trainees have engaged in projects such as working in the universities' technology-transfer offices, teaming up with professional science writers and lobbying state legislature.

Patricia Labosky, who is the programme leader for BEST at the NIH in Bethesda, Maryland, says that neither career training nor internships are an entirely new endeavour for graduate schools. However, she adds that the agency's approach is innovative — it could transform graduate internships and career training from an ad hoc approach to a more-systematic, data-driven one.

The best of the BEST projects will probably be adopted by more universities once the pilot expires in 2018, says Kathy Gould, a biologist who runs the BEST programme at Vanderbilt University in Nashville, Tennessee. "What we're seeing is an experiment in progress," she says. Labosky predicts that market forces could speed up adoption because BEST universities use the initiative as a student-recruiting tool.

FACULTY FIRST

Faculty-member buy-in to the programme is essential because many principal investigators (PIs) look at students' or postdocs' time away from the lab as detrimental — especially if those young scientists are funded under the PI's grant. But some have come around, says Ambika Mathur, who teaches paediatrics at Wayne State University in Detroit, Michigan, and is dean of its graduate school. When her institution surveyed science faculty members to gauge reaction to off-site internships for trainees, most PIs said that they were favourably or very favourably inclined.

One solution to objections from PIs is to fund trainees who are on internships from departmental money or training grants, which are not tied explicitly to research outcomes. Some BEST universities reimburse PIs for the time that their trainees spend away from the lab. But such a change does not happen quickly or easily. "Shifting trainees off research grants is going to take a long time to accomplish," says Nael McCarty, who manages Emory University's BEST programme in Atlanta, Georgia. And tweaking funding ►

► sources alone will not address PIs' need for young scientists to get experiments done. "Our entire careers are resting on the backs of our trainees," he points out.

SPLIT THE DIFFERENCE

Emory is quelling some faculty members' objections by casting internships as separate, independent projects with flexible hours. The university also sets a low bar on the schedule and time requirements for internships. BEST advisers first sign off on their students and postdocs joining the programme, with the understanding that the trainees will spend at least 50 hours each semester or during their summer academic break away from the lab. Those hours could be compressed into a month or even a week, or they could be spread out over a semester or a year. Once trainees find potential placements, they negotiate their internship schedules with their PIs.

This negotiation was a fraught process for Chelsey Ruppensburg. As a PhD student, also

at Emory, she created a public-policy internship for herself in Emory's office of government and community affairs that required 20 hours a week at the Georgia State Capitol, where she advocated to the state legislature on behalf of the university.

"It started with a frank conversation with my PI — that I would be out of his lab during traditional hours," she says. "I was going to do everything on my end to make sure this wouldn't harm my PI and my work in his lab." They agreed that she could spend two or three days each week at the Capitol from January to April during Georgia's legislative session. She made up lab hours at nights and on weekends, hustling to finish her doctoral dissertation on cell biology while learning the ropes of advocacy. The extra work paid off: Ruppensburg started a post last month as a fundraising staffer in the political campaign for US senator Johnny Isakson.

The PI is not the only one who can pose an obstacle to crafting a useful internship.

Finding employers who offer the kind of flexibility that graduate- and postdoc-level interns require can be tricky, says Gould. She advises trainees to think carefully about what they want from an internship, and she asks prospective partners to plan projects that are

"It started with a frank conversation with my PI — that I would be out of his lab during traditional hours."

appropriate for a graduate student or postdoc's skill level and training needs. "Some people in industry, on the surface, are very enthusiastic," she says. But they often prove unable to write a job descrip-

tion for a PhD-level scientist.

To sidestep these challenges, some BEST trainees create internships through their own university, like Ruppensburg did. Most universities can offer opportunities in areas such as science writing and intellectual-property management. Many BEST schools have found it easier to place students in internships on campus than offsite, especially if they are not located in or near a major technology hub, says Labosky.

For example, the BEST programme at New York University (NYU) works with the university's technology-transfer office. A formal internship charges the trainee with drafting a business plan around a particular piece of technology. A less-formal option pairs graduate students and postdocs with companies to write marketing summaries, gather competitive intelligence and perform outreach. Neither option requires fixed hours, but before graduate students or postdocs can participate, they must complete a technology-commercialization course. At this point, about 40 students and postdocs have interned with NYU's tech-transfer office in one of these capacities.

Other approaches to internships will arise as the BEST programme adapts to the needs of trainees, PIs and internship sponsors, says Keith Micoli, director of NYU's medical school and co-PI of its BEST grant. Short internships can be effective if they help a trainee to choose or rule out a particular career pathway, he adds. By encouraging trainees to explore different options — even through simple things such as job shadowing — these programmes could help to ease the bottleneck of graduate students and postdocs who do not know what they want to do once they complete their programme, he adds (see "The rule of threes"). "One of the most frustrating things I see," he says, "is graduate students who complete their PhDs and say, 'I suppose I should do a postdoc and figure out what I need to do.'"

As the BEST model expands, Micoli and others hope to hear that less often. ■

Paul Smaglik is a freelance writer in Milwaukee, Wisconsin.

BEST CALCULATIONS

The rule of threes

Douglas White knows how to work the numbers — and how to make them work for him. About halfway through his doctoral programme in biomedical engineering at the Georgia Institute of Technology (Georgia Tech) in Atlanta, he realized that the odds of landing a tenure-track position were not on his side. To prepare for a career with better chances of employment, White turned to the US National Institutes of Health's Broadening Experience in Scientific Training (BEST) programme.

Today, thanks to three internships, White works as a project manager at Takeda Pharmaceuticals, a Japan-based drugmaker with a presence in Atlanta, Georgia. "People always talk about how internships lead to job opportunities, but I didn't believe it," he says.

BEST offered flexible internship opportunities that allowed him to experience different paths over about 18 months. First, he completed a writing apprenticeship sponsored by the philanthropic W. M. Keck Foundation in Los Angeles, California, which paired him with two professional science writers, put him in touch with about a dozen science-writing students around the country and sent him to a scientific conference where he reported on its proceedings and was critiqued by professionals.

He learned that he loves writing, but did not want to pursue it as a career. Equally important, he learned how to tailor his message for people from different

backgrounds and with various levels of scientific knowledge.

Next, he decided to look into the government sector, and the Atlanta BEST office came through with an application for an internship at the US Defense Forensic Science Center, about 24 kilometres from Georgia Tech in Forest Park.

During the interview, White was excited to learn that the centre used a technology he was interested in. However, after two months, he realized that the hours out of the lab were cutting into his dissertation research and that he was spending little time on the aspects of the work that were most relevant to his training and research needs.

The experience gave him a crash course in negotiating an early exit. He also learned that he needed to consider work-life balance, so when he landed an interview for his next internship with Takeda, he told them that he was scheduled to defend his dissertation in a few months and was getting married a week after that. He won the internship, defended his thesis and left for his honeymoon — and when he returned, his department had been restructured and his new boss offered him a job as a project manager.

White says that the three internships changed his life by allowing him to explore multiple career options in a relatively short amount of time — and landed him a full-time job in the process. "I would not be where I am if it wasn't for the BEST programme," he says. **P.S.**

TURNING POINT

Anja Rammig

Ecosystem modeller Anja Rammig started in June as an assistant professor at Germany's Technical University of Munich (TUM), which in 2012 adopted a tenure-track scheme.

Was there a pivotal moment in your career?

I was working on my undergraduate degree in zoology, and I was so excited to learn it was possible to do computer simulations of ecosystems that I almost changed my course of study. My adviser suggested that instead, I pursue a PhD thesis focused on computer modelling. It was the most important move in my scientific life.

How did you come to specialize in forests?

During my PhD programme at the Swiss Federal Institute of Technology in Zurich, I worked with researchers at the Institute for Snow and Avalanche Research. Switzerland uses forests as avalanche protection, and researchers had collected data after a strong windstorm in 1990 that had killed many trees in the country. They wanted to learn how long it would take for the forest to regenerate. It was my first experience with modelling, and it convinced me that I wanted to continue this type of work but with a focus on global problems.

What was your seven-year experience at the Potsdam Institute for Climate Impact Research (PIK) in Germany like?

Working at this world-renowned institute was my introduction to the big world of science policy. In the first month, I began estimating the large-scale die-off of the Amazon rainforest as a result of climate change. I gave presentations at the World Bank and other organizations.

Did you get any coaching on communication?

I had discussions with colleagues about how to communicate the science, but I got no specific coaching. For me, science communication was very new, and in my first two years at PIK, it was difficult to do. But I learned what audiences expected and what level of information worked best. It was the culture of the institute to learn by watching more-experienced colleagues. I determined that it is really important to read a lot to prepare for questions from scientists, stakeholders or politicians — and to know the Intergovernmental Panel on Climate Change reports almost by heart.

How is it being a woman in such a male-dominated field?

PIK was trying to increase the number of women in high positions; they were very fair when it came to parental leave and work-life



balance. I was in the group at PIK with the highest percentage of women, incidence of maternity leave and employees with children. My last three years there, I had a female boss and worked with five other women. Last year, I applied for a position in Germany that aimed to attract female applicants. I was pregnant at the time, and the date for my presentation was my due date. Ironically, they wouldn't move my presentation date, so they didn't consider me.

Can you describe TUM's tenure-track scheme?

The criteria for how you will be evaluated — on research, teaching and public engagement — are clearly spelled out. Research criteria include developing methodologies and concepts, securing external funding and showing that you are building an international reputation. It's not like the impression you may get at other institutions: that is, that the process is not transparent or lacks defined criteria. Tenure-track professorships in all disciplines have all the same criteria.

What landed you the position at TUM?

I really wanted my own research group to keep studying the Amazon's ecophysiology and how it might change. I'm interested in modelling it with data from experimental studies. Fortunately, while I was at PIK, I established a huge network of collaborations and connections, especially with Brazilian scientists. I am on the scientific committee of a large collaboration to build a big experiment in the Amazon rainforest that will test the impact of increasing carbon dioxide. I think my connections helped me, even though competition for the job at TUM was high. ■

INTERVIEW BY VIRGINIA GEWIN

This interview has been edited for length and clarity.

TIME FLIES

The catch of a lifetime.

BY CARIE JUETTNER

“Dang! I almost had it.”
“No, you didn’t. You were a mile away.”

“Was not. It was a good one too. It was tiny.”

“That’s a myth, you know,” I said, swiping at a medium-sized green one with my baseball cap. “Size doesn’t matter.”

“Yeah it does, Kat.”

Jeremy put his net away and got out a pair of chopsticks. I rolled my eyes. “The small ones are always worth more. Haven’t you ever noticed that the big slow ones that are so easy to catch are never worth more than an hour or two?”

“So? I’m starting to think there are no valuable ones anymore. I think they went extinct or something.”

“No, they’re out there. My friend Jon caught a 17-year one once!”

“Oh yeah? Did you actually see it?” I swiped my foot through the tall grass, trying to stir something up.

“No, but I believe him.”

“Seventeen years, huh? So does Jon look different now?”

A small iridescent blue one flew by. Jeremy made a flailing jab at it and missed. “No. It got away before he could swallow it.”

“Pshh. Then he was lying.” I made an attempt at the blue one, but it was already out of my reach. “There’s nothing bigger than a year out there anymore. Sometimes I think there’s nothing bigger than a week.”

“If you really think that, then why do you even catch?”

I shrugged. “Every little bit helps.”

We were quiet for a while. A thick-bodied brown buzzer lifted clumsily out of the grass and I grabbed it without even using my cap. I flipped it over. Three hours. Turning my back to Jeremy, I popped the whole thing in my mouth at once. I bit down and the buzzing stopped as the earthy metallic taste flooded my tongue. I



swallowed and waited, eyes closed, for the sensation I knew would come. I had to concentrate or I would miss it. A quick burst of energy and warmth passed through me, and then it was gone. I opened my eyes.

“You know,” Jeremy said, “my grandfather hated catching. He wouldn’t let me do it when he was around. He said we should be happy with the time we’re born with.”

I scoffed. “And where is he now?”

“He’s dead.”

I nodded. It was starting to get dark. There’d be more coming soon, but they’d be harder to see.

“How long do you think you’ll do it?” I asked.

“Do what?”

“Catch.”

Jeremy shrugged. “Not long. I just want to build up little more, you know? Bank it. Then I’ll stop. I mean, I’m not going to do it for ever.”

“Me neither.” I knew it was rude to ask, but I couldn’t help myself. “How much do you have? You know, in the bank?”

Jeremy examined his shoes. “A good amount. I’m comfortable.”

I nodded.

“Yeah,” he said. “Just one more good catch, and I’ll hang up my net.” He smiled at me. “And my chopsticks.”

I smiled back.

I heard a high-pitched buzzing behind me. I turned around and saw a minuscule flyer heading right for us. I pulled my cap back, but Jeremy was faster. He dropped his chopsticks and clasped his bare hands around it, trapping it inside. “I got it!” he yelled. “I actually got it!” He whooped and jumped up and down. “It’s the smallest one I’ve ever seen!”

“So look at it already. What is it?” I crossed my arms to keep them from shaking. *What if the myth was true after all?*

Jeremy moved his hands carefully until he had the tiny creature pinned firmly between his fingers. He flipped it over and peered at the underside. For just a moment, his face fell. Then his smile was back. He popped the thing in his mouth, swallowed loudly and stood still for a long time. I let him have his moment, even though it was killing me to wait. *What was it? Ten years? Twenty?* Finally, he let out a deep sigh and looked at me.

“How much?” I asked.

“Six.”

“Six years?”

He nodded. My arms relaxed a little.

“That’s a pretty good catch,” I said.

He nodded again, smiling slightly. “Yeah.”

“So,” I said, “since you’re not going to be needing that net ...?” I held my hand out towards him.

A hum filled the air as several large flyers and a few medium-sized ones emerged from the grass. Jeremy shuffled his feet, his fingers gripping the net. “Actually ... I think ... maybe I’ll try for just one more.” ■

➔ NATURE.COM

Follow Futures:

@NatureFutures

go.nature.com/mtoodm

Carie Juettner is a writer in Austin, Texas. Her fiction has appeared in Hello Horror, Dark Moon Digest, Microhorror and Writers Weekly, among other places. Follow her blog at www.cariejuettner.com.